Systèmes de compréhension et de traduction de la parole : vers une approche unifiée dans le cadre de la portabilité multilingue des systèmes de dialogue

Bassam Jabaian

ACADÉMIE D'AIX-MARSEILLE
UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE

THÈSE

présentée à l'Université d'Avignon et des Pays de Vaucluse
pour obtenir le diplôme de DOCTORAT

SPÉCIALITÉ : Informatique

École Doctorale 536 « Sciences et Agronomie »
Laboratoire d'Informatique (EA 4128)

Systèmes de compréhension et de traduction de la parole : vers une approche unifiée dans le cadre de la portabilité multilingue des systèmes de dialogue

par
Bassam Jabaian

Soutenue publiquement le ? ? ? 2012 devant un jury composé de :

M.	François Yvon	Professeur, LIMSI, Paris	Rapporteur
M.	Wolfgang Minker	Professeur, ITDS, Ulm	Rapporteur
M^{me}	Géraldine DAMNATI	Ingénieur de recherche, Orange Labs, Lannion	Examinateur
M.	Alex Allauzen	Maitre de Conférences, LIMSI, Paris	Examinateur
M.	Laurent Besacier	Professeur, LIG, Grenoble	Directeur de thèse
M.	Fabrice Lefèvre	Professeur, LIA, Avignon	Directeur de thèse

Laboratoire d'Informatique d'Avignon - CERI
Centre d'Enseignement et de Recherche en Informatique

Résumé

La généralisation de l'usage des systèmes de dialogue homme-machine accroît la nécessité du développement rapide des différents composants de ces systèmes. Les systèmes de dialogue peuvent être conçus pour différents domaines d'application et dans des langues différentes. La nécessité d'une production rapide pour de nouvelles langues reste un problème ouvert et crucial auquel il est nécessaire d'apporter des solutions efficaces.

Nos travaux s'intéressent particulièrement au module de compréhension de la parole et proposent des approches pour la portabilité rapide peu coûteuse de ce module. Les méthodes statistiques ont montré de bonnes performances pour concevoir les modules de compréhension de la parole pour l'étiquetage sémantique de tours de dialogue. Cependant ces méthodes nécessitent de larges corpus pour être apprises. La collecte de ces corpus est aussi coûteuse en temps et en expertise humaine.

Dans cette thèse, nous proposons plusieurs approches pour porter un système de compréhension d'une langue vers une autre en utilisant les techniques de la traduction automatique. Les premiers travaux consistent à appliquer la traduction automatique à plusieurs niveaux du processus de portabilité du système de compréhension afin de réduire le coût lié à production de nouvelles données d'apprentissage. Les résultats expérimentaux montrent que l'utilisation de la traduction automatique permet d'obtenir des systèmes performant avec un minimum de contribution humaine.

Cette thèse traite donc à la fois de la traduction automatique et de la compréhension de la parole. Nous avons effectué une comparaison approfondie entre les méthodes utilisées pour chacune des tâches et nous avons proposé un décodage conjoint basé sur une méthode discriminante qui à la fois traduit une phrase et lui attribue ses étiquettes sémantiques. Ce décodage est obtenu par une approche à base de graphe qui permet de composer un graphe de traduction avec un graphe de compréhension. Cette représentation peut être généralisée pour permettre des transmissions d'informations riches entre les composants du système de dialogue.

Table des matières

1 Introduction **9**
- 1.1 Contexte général . . . 10
- 1.2 Objet de la thèse . . . 11
- 1.3 Projet PORT-MEDIA . . . 13
- 1.4 Organisation du document . . . 14

I Cadre théorique **17**

2 La compréhension automatique de la parole **19**
- 2.1 Introduction . . . 20
- 2.2 Le système de compréhension de la parole . . . 22
- 2.3 Représentations sémantiques pour la compréhension . . . 23
- 2.4 Approches pour la compréhension de la parole . . . 26
  - 2.4.1 Approches linguistiques . . . 26
  - 2.4.2 Approches issues de l'apprentissage automatique . . . 28
    - 2.4.2.1 Les modèles Markoviens . . . 29
    - 2.4.2.2 Approche par traduction automatique (SMT) . . . 32
    - 2.4.2.3 Les transducteurs à états finis (FST) . . . 32
    - 2.4.2.4 Les machines à vecteur de support (SVM) . . . 33
- 2.5 Evaluation des systèmes de compréhension de la parole . . . 34
- 2.6 Conclusion . . . 35

3 La traduction automatique **37**
- 3.1 Introduction . . . 38
- 3.2 Architectures des systèmes de traduction . . . 39
  - 3.2.1 Architectures linguistiques . . . 39
  - 3.2.2 Architectures computationnelles . . . 40
- 3.3 La traduction automatique probabiliste . . . 41
  - 3.3.1 Modèle de langage . . . 42
  - 3.3.2 Modèle de traduction . . . 44
    - 3.3.2.1 Traduction à base de mots . . . 45
    - 3.3.2.2 Traduction à base de segments . . . 48
  - 3.3.3 Modèle log-linéaire . . . 49
  - 3.3.4 Décodage . . . 51

3.3.5 Approche hiérarchique pour la traduction automatique 53
3.4 Outils pour la traduction automatique probabiliste 55
3.5 Evaluation des systèmes de traduction 56
3.6 Conclusion 57

II La portabilité multilingue d'un système de compréhension automatique de la parole 59

4 La portabilité d'un système de compréhension de la parole 61
4.1 Introduction 62
4.2 La portabilité des systèmes de dialogue 62
4.2.1 La portabilité des systèmes de reconnaissance automatique de la parole 63
4.2.2 La portabilité des systèmes de compréhension 64
4.3 Nos approches pour la portabilité multilingue d'un système de compréhension 66
4.3.1 La portabilité au niveau du décodage (TestOnSource) 67
4.3.2 La portabilité au niveau de l'apprentissage (TrainOnTarget) . . . 68
4.4 La portabilité de l'annotation sémantique 69
4.4.1 Alignement direct (non-supervisé) 69
4.4.2 Alignement indirect (semi-supervisé) 70
4.4.3 Alignement obtenu pendant la traduction 71
4.5 Accroître la robustesse du système de compréhension aux erreurs de traduction 73
4.5.1 Apprentissage sur des données bruitées (SCTD) 74
4.5.2 Post-édition statistique (SPE) 75
4.6 Conclusion 76

5 Portabilité : expériences et résultats 77
5.1 Introduction 78
5.2 Matériau expérimental 78
5.2.1 Le corpus MEDIA 78
5.2.2 Les métriques d'évaluation 81
5.2.3 Les systèmes de traduction 81
5.3 Evaluation des approches proposées pour la portabilité 83
5.3.1 Les modèles de référence 83
5.3.2 Evaluation sur la totalité des données 84
5.3.3 Evaluation sur des données partielles 86
5.3.4 Evaluation des méthodes robustes aux erreurs de traduction . . . 87
5.3.5 Combinaison 88
5.4 Validation des approches de portabilité proposées 89
5.4.1 Validation en utilisant des traductions en ligne 89
5.4.2 Validation sur une autre langue (arabe) 91
5.4.3 Pré-annotation automatique 93
5.5 Conclusion 96

III Approches conjointes pour la traduction et la compréhension **99**

6 Génératif vs. discriminant pour la traduction et la compréhension **101**
6.1 Introduction . . . 102
6.2 Méthode de traduction pour la compréhension . . . 103
6.2.1 Adaptation des méthodes . . . 104
6.2.2 Application à la portabilité multilingue . . . 106
6.3 Méthode de compréhension pour la traduction . . . 106
6.3.1 Modèle du LIMSI (FST/CRF) . . . 107
6.4 Décodage conjoint pour la traduction et la compréhension, cas de la méthode de portabilité TestOnSource . . . 112
6.5 Conclusion . . . 114

7 Approches conjointes : expériences et résultats **117**
7.1 Introduction . . . 118
7.2 Evaluation des systèmes de traduction à base de segments pour une tâche de compréhension . . . 118
7.2.1 Evaluation du système français . . . 118
7.2.2 Evaluation du système italien . . . 120
7.3 Evaluation des systèmes de traduction à base de CRFs . . . 121
7.3.1 Evaluation du système de traduction français vers italien . . . 122
7.3.2 Evaluation du système de traduction italien vers français . . . 123
7.4 Evaluation des systèmes de traduction selon l'approche FST/CRF . . . 124
7.4.1 Evaluation pour une tâche de traduction . . . 124
7.4.2 Evaluation pour une tâche de compréhension . . . 126
7.5 Décodage conjoint dans le cas d'un scénario de portabilité du français vers l'italien d'un système de compréhension (TestOnSource) . . . 126
7.6 Conclusion . . . 129

8 Conclusions et Perspectives **131**
8.1 Conclusion . . . 132
8.2 Perspectives . . . 134

Liste des illustrations **137**

Liste des tableaux **139**

Bibliographie **141**

Bibliographie personnelle **155**

Chapitre 1

Introduction

Sommaire

1.1 Contexte général 10
1.2 Objet de la thèse 11
1.3 Projet PORT-MEDIA 13
1.4 Organisation du document 14

La fin du vingtième siècle a vécu un grand changement avec l'accroissement des possibilités de communication entre l'homme et la machine. L'évolution des différents modules de traitement automatique du langage naturel a permis d'accélérer l'évolution vers une société numérique organisé autour des fonctionnalités offertes par les ordinateurs. Cependant l'interaction homme-machine utilisant la parole reste un défi de grande importance pour une large catégorie de personnes pour lesquelles c'est le seul moyen pratique d'accès à l'information. Les systèmes de dialogue oraux permettent à un utilisateur non spécialiste de communiquer avec un système complexe pour accéder facilement à l'information.

1.1 Contexte général

L'architecture d'un système de dialogue homme-machine varie selon le niveau d'interaction dans ce système. Les systèmes de dialogue peuvent être classés dans deux catégories principales : les systèmes de dialogue guidé et les systèmes de dialogue à initiative mixte.

Dans la première catégorie, le système de dialogue pose des questions précises à l'utilisateur et attend une réponse directe et simple. Ces systèmes, peu complexes, nécessitent peu d'analyse sémantique et peuvent être conçus facilement car ils sont généralement déterminés par une liste de grammaires prédéfinies. Dans les systèmes à initiative mixte l'utilisateur et le système peuvent intervenir à n'importe quel moment du dialogue pour demander une information ou préciser une requête. Cette caractéristique rend le système moins artificiel mais plus difficile à construire. L'analyse sémantique des requêtes de l'utilisateur est une étape délicate et indispensable dans ce type de système.

Un système de dialogue interactif est composé de plusieurs modules qui fonctionnent séquentiellement (voir la figure 2.1). Le signal audio de l'énoncé utilisateur est transcrit automatiquement par un système de reconnaissance de la parole. Cette transcription est ensuite transférée au système de compréhension de la parole qui attribue des étiquettes sémantiques à cette phrase. Ces étiquettes représentent le sens contenu dans la requête. Le gestionnaire de dialogue prend des décisions sur la procédure à suivre dans le dialogue en se basant sur les informations qu'il récupère sur l'état actuel du dialogue et en utilisant des sources extérieures (base de données, web...). Les actions décidées par le gestionnaire de dialogue sont transmises à l'utilisateur à l'aide d'un système de génération de textes et d'un système de synthèse de la parole.

Un grand défi lors de la construction d'un système de dialogue est non seulement de pouvoir construire les meilleurs composants du système, mais aussi de réduire le temps nécessaire pour cette production.

Dans les systèmes de dialogue actuels, la plupart des composants reposent sur des approches probabilistes, qui forment une alternative efficace au modèles précédents à base de règles. La construction de ces modèles nécessite un nombre important de données d'apprentissage sur lesquelles sont appris les modèles.

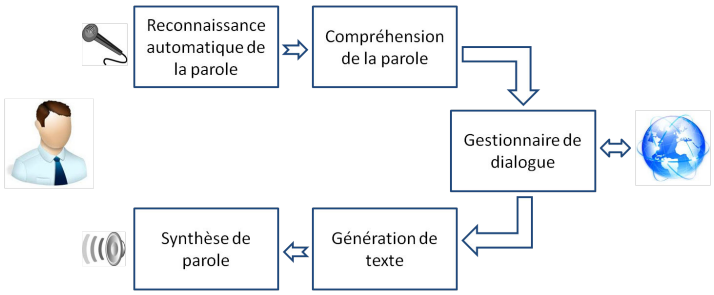

FIGURE *1.1* – *Architecture générale d'un système de dialogue.*

(Gao et al., 2005) ont montré que la partie la plus coûteuse en temps pour construire un nouveau système de dialogue est de collecter, de transcrire et d'annoter des données pour le développement du module de compréhension. Afin de pouvoir minimiser le temps nécessaire pour la production d'un système de dialogue, des approches pour faciliter la production de telles données ou d'une production rapide d'un tel module sont donc indispensables.

1.2 Objet de la thèse

Dans cette thèse nous nous intéressons à la portabilité multilingue d'un système de compréhension de la parole afin de minimiser le temps nécessaire pour la production d'un nouveau système dans une nouvelle langue.

Le principe de la portabilité d'un système de compréhension de la parole consiste à utiliser un système existant (et les sources qui ont servi à le construire) afin de pouvoir obtenir un nouveau système avec un coût réduit en temps et en expertise humaine. Cette portabilité peut être appliquée pour obtenir de nouveaux systèmes pour des langues ou domaines différents.

A la suite de travaux precurseurs dans les années 2000 (Minker, 1998) le multilinguisme des systèmes de compréhension est un sujet (ré-)abordé de plus en plus ces dernières années (Tur et al., 2003; Fung and Schultz, 2008; Lefèvre et al., 2010) . Le renouveau de la recherche dans la portabilité multilingue d'un système de compréhension de la parole peut être expliquée par deux facteurs importants :

- l'évolution du paradigme dominant des systèmes de compréhension vers des approches statistiques permettant de construire des modèles à partir de corpus étiqueté ;
- le fait que les méthodes de traduction automatique sont désormais assez performantes et beaucoup plus faciles à développer et à adapter à un besoin précis.

Cette thèse est à l'intersection de deux domaines scientifiques : la compréhension de la parole et la traduction automatique. Plus précisément dans nos travaux nous cherchons à utiliser les techniques de la traduction automatique afin de porter un système de compréhension existant vers une nouvelle langue.

Les systèmes de compréhension utilisés récemment sont des systèmes statistiques basés sur des corpus d'apprentissage annotés sémantiquement. Ces corpus forment une représentation sémantique d'un domaine précis. La portabilité de ces systèmes consiste à pouvoir porter le sens contenu dans ces corpus pour pouvoir étiqueter des énoncés dans une nouvelle langue.

Cette portabilité peut être obtenue soit par la traduction de ces corpus (avec un moyen efficace de projeter l'annotation sémantique dans le corpus traduit) soit en apprenant des modèles sur ces corpus et traduisant les énoncés de l'utilisateur vers la langue sur laquelle ces modèles sont appris, ceci afin d'étiqueter la phrase traduite avec les modèles appris.

Dans les deux cas, un système de traduction automatique performant est nécessaire pour obtenir des corpus traduits de bonne qualité. Ces traductions peuvent être obtenues par un système de traduction générique (les systèmes disponibles en ligne par exemple) ou par un système de traduction spécialisé qui pourra être appris sur un corpus bilingue collecté spécialement.

Les travaux réalisés dans cette thèse sont motivés par la disponibilité du corpus de dialogue MEDIA et d'une traduction manuelle d'une sous-partie de ce corpus vers l'italien. MEDIA (Bonneau-Maynard et al., 2005) est un corpus de dialogue français pour des informations touristiques et réservation d'hôtel, collecté avec les techniques de magicien d'OZ pour simuler un dialogue homme-machine. Ce corpus est composé de plus d'un millier de dialogues étiquetés par une centaine de concepts qui représentent la sémantique du domaine.

Nous prenons l'avantage d'avoir une sous-partie de ce corpus traduite manuellement afin d'avoir un corpus parallèle spécialisé pour apprendre des systèmes de traduction statistiques en utilisant des outils libres (Koehn et al., 2007). Ces systèmes seront utilisés ensuite pour la mise en place de nos propositions de portabilité.

Les travaux de cette thèse suivent deux axes principaux. Nous consacrons la première partie à la portabilité multilingue d'un système de compréhension de la parole. Pour cela nous proposons plusieurs méthodes qui permettent de porter un système de compréhension de la parole vers une nouvelle langue en se basant sur les techniques de la traduction automatique.

Ces méthodes peuvent être classées en deux catégories qui diffèrent par le niveau auquel la portabilité est appliquée. La première approche cherche à étiqueter les phrases de la nouvelle langue sans forcément chercher à avoir un système de compréhension dans cette langue. Pour cela nous proposons d'utiliser la traduction automatique pour traduire les entrées de la nouvelle langue afin de pouvoir les étiqueter par un système existant.

La deuxième approche cherche à apprendre un nouveau système dans la nouvelle langue. Un tel système nécessite un corpus d'apprentissage dans cette langue. Pour cela nous proposons d'utiliser la traduction automatique pour traduire les données d'apprentissage et nous proposons plusieurs techniques pour porter l'annotation du corpus existant vers le corpus traduit.

Dans un second temps, nous cherchons à démontrer les relations entre les différentes approches statistiques utilisées pour la compréhension et pour la traduction afin de pouvoir les combiner et d'obtenir une approche conjointe pour la portabilité du système de compréhension. La problématique qui se pose ici est liée à la difficulté de trouver un modèle homogène pour les deux approches.

L'approche qui donne la meilleure performance en traduction n'est pas celle qui donne la meilleure performance en compréhension malgré toutes les adaptations qu'on peut appliquer sur l'une ou l'autre des approches pour l'appliquer efficacement à la nouvelle tâche. Dans cette thèse nous proposons une approche à base de graphes qui permet de combiner ces deux approches et d'obtenir un décodage conjoint qui à la fois traduit des énoncés et produit leur annotation sémantique.

1.3 Projet PORT-MEDIA

Cette thèse s'inscrit dans le cadre du projet ANR PORT-MEDIA [1], financé dans le cadre de l'appel à projet Contenu et Interactions. Les travaux de recherche ont été effectués en collaboration entre le Laboratoire Informatique de Grenoble (LIG) et le Laboratoire Informatique d'Avignon (LIA).

Aujourd'hui, plusieurs applications commerciales sont fondées sur la reconnaissance vocale. La qualité de l'interaction homme-machine est encore loin d'être agréable et efficace. Un bon moyen d'améliorer l'utilisabilité et l'acceptabilité des systèmes de dialogue automatiques est d'augmenter le niveau d'intelligence des systèmes automatiques, notamment en améliorant la partie en charge de la compréhension du langage parlé.

Dans cette perspective, le projet PORT-MEDIA est une suite naturelle de la campagne d'évaluation de systèmes de compréhension EVALDA/MEDIA.

Le projet MEDIA a donné aux laboratoires publics français et aux industriels intéressés par la compréhension de la parole dans le cadre d'un système de dialogue homme-machine, une plateforme commune pour l'évaluation de leurs systèmes de compréhension, à la fois avec ou sans contexte de dialogue.

Ces dernières années ont vu émerger des approches statistiques pour la compréhension automatique de la parole, élément clé des systèmes de dialogue oraux, qui extrait le sens des énoncés utilisateurs. L'objet du présent projet PORT-MEDIA est d'enrichir le corpus MEDIA avec trois aspects complémentaires de grande importance dans les systèmes de dialogue :

– **La robustesse :** intégration du module de la reconnaissance automatique de la parole dans le processus de compréhension.

1. http ://www.port-media.org

– **La portabilité à travers les domaines et les langues :** évaluation de la généricité et la capacité d'adaptation des approches mises en œuvre dans les systèmes de compréhension. Cela est fait en confrontant les modèles à des nouvelles données produites pour une autre tâche ou dans une autre langue, puis aussi en exploitant des données annotées.

– **Des structures riches de haut niveau pour la représentation des connaissances sémantiques :** des règles solides pour les unités sémantiques de base (concepts) ont été proposées et testées au cours du premier projet. La représentation sémantique de MEDIA est enrichie par une nouvelle représentation sémantique de haut niveau permettant une représentation complète de la composition sémantique à l'intérieur de tours utilisateurs.

Les objectives du projet PORT-MEDIA peuvent être résumés de la manière suivante :

– **Production de métadonnées :** cela consiste à obtenir une transcription automatique du corpus MEDIA avec plusieurs niveaux de qualité. Le projet cherche aussi à définir un nouveau formalisme pour la représentation sémantique de haut-niveau basé sur le langage MMIL.

– **Production rapide de nouvelles données :** cette production est basée sur une pré-transcription automatique des données du nouveau domaine et aussi sur une pré-annotation sémantique automatique des données nouvelles (langue et domaine). La figure 1.2 montre une vue globale sur le processus de production rapide de données dans le contexte du projet PORT-MEDIA.

En ligne avec les objective du projet, cette thèse propose des méthodes de portabilité d'un système de compréhension vers une nouvelle langue. Ces méthodes seront utilisées pour fournir une pré-annotation sémantique pour le processus de production de nouvelles données.

1.4 Organisation du document

Ce document est organisé en trois parties principales : dans la mesure où cette thèse est à l'intersection entre la compréhension automatique de la parole et la traduction automatique, la première partie de ce document présente un état de l'art de ces deux domaines en présentant les différents paradigmes dominants et les méthodes utilisées pour les mettre en œuvre.

Après une brève présentation des différents composants du système de dialogue, nous présentons dans le chapitre 2 un état de l'art de la compréhension de la parole. Nous définissons les notions de représentation sémantique et nous présentons les différentes approches utilisées pour concevoir un système de compréhension automatique ainsi que les mesures utilisées pour évaluer les performances de ce module.

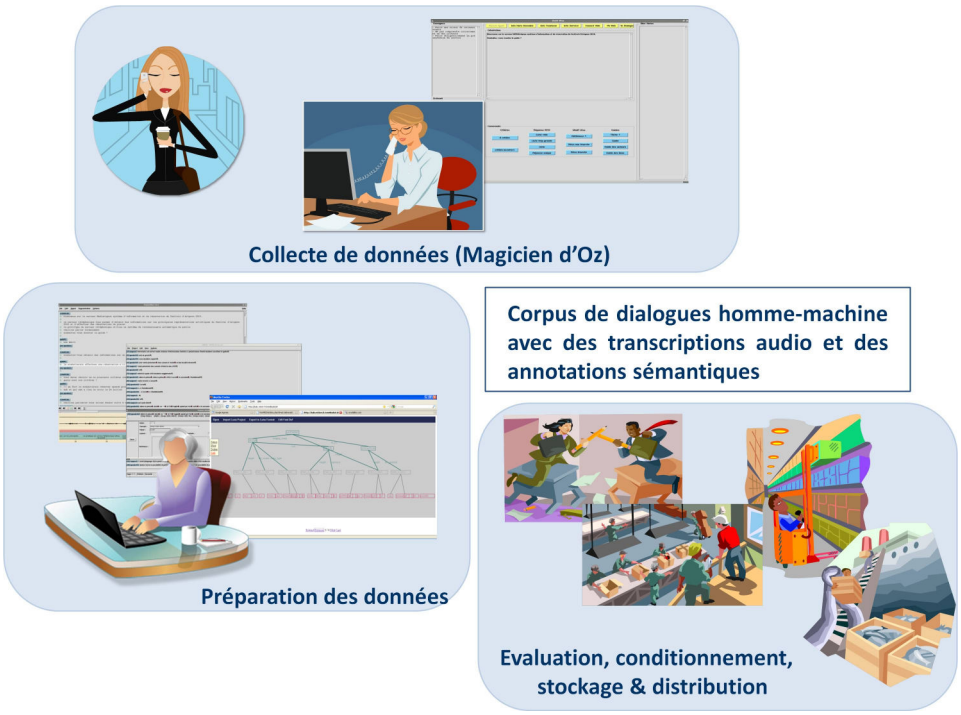

FIGURE 1.2 – *Illustration du projet ANR PORT-MEDIA : le processus de production de nouvelles données.*

Dans le chapitre 3, nous présentons un état de l'art de la traduction automatique en mettant l'accent sur la traduction automatique probabiliste. Nous présentons les différents composants de l'approche de traduction à base de segments (modèle de langage, modèle de traduction et modèle de réordonnancement), le décodage en traduction automatique ainsi que les mesures d'évaluation utilisées.

La deuxième partie de ce document porte sur la portabilité multilingue du système de dialogue et notamment sur la portabilité d'un système de compréhension statistique. Dans le chapitre 4 nous présentons des travaux dédiés à la portabilité ainsi que nos propositions pour une portabilité multilingue de la compréhension. Une étude expérimentale et des évaluations sur les différentes approches proposées dans cette thèse sont présentées dans le chapitre 5.

La troisième partie est dédiée à la comparaison entre les approches génératives et les approches discriminantes aussi bien pour la compréhension de la parole que pour la traduction automatique. Dans le chapitre 6, nous analysons les différentes approches et nous présentons nos propositions pour adapter ces approches à une tâche ou à une autre. Nous proposons aussi dans ce chapitre une approche qui permet d'obtenir un décodage conjoint entre la traduction et la compréhension.

Nos propositions d'adaptation de systèmes et notre proposition de décodage conjoint sont évaluées dans le chapitre 7. Enfin le chapitre 8 conclut ce travail et présente des perspectives

Première partie

Cadre théorique

Chapitre 2

La compréhension automatique de la parole

Sommaire

2.1 Introduction 20
2.2 Le système de compréhension de la parole 22
2.3 Représentations sémantiques pour la compréhension 23
2.4 Approches pour la compréhension de la parole 26
2.4.1 Approches linguistiques 26
2.4.2 Approches issues de l'apprentissage automatique 28
2.4.2.1 Les modèles Markoviens 29
2.4.2.2 Approche par traduction automatique (SMT) 32
2.4.2.3 Les transducteurs à états finis (FST) 32
2.4.2.4 Les machines à vecteur de support (SVM) 33
2.5 Evaluation des systèmes de compréhension de la parole 34
2.6 Conclusion 35

2.1 Introduction

L'utilisation des systèmes de dialogue homme-machine commence à se répandre dans la vie quotidienne. La mise à disposition de dispositifs d'interrogation vocale dans les dernières versions de smartphones (avec Google Voice sous Android ou Siri sous iOS) constitue un accélérateur certain de cette tendance. Le système de dialogue permet à l'utilisateur une communication orale avec la machine afin de résoudre un problème ou de rechercher une information.

L'architecture de ces systèmes est variable selon le niveau d'interaction. Dans certains systèmes, les dialogues sont guidés (e.g. répondre par oui ou par non), alors que dans d'autres systèmes les dialogues sont plus libres (e.g. quelle information désirez-vous avoir ?) et donc l'architecture du système est plus complexe. La figure 2.1 représente l'architecture d'un système de dialogue interactif. Ce système est composé de plusieurs modules qui fonctionnent séquentiellement.

Un panorama plus large sur les systèmes de dialogue homme-machine pourra être trouvé dans (Minker and Bennacef, 2004) ou encore (McTear, 2004).

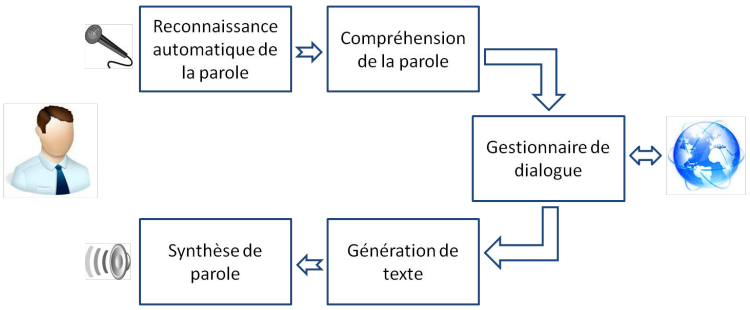

FIGURE 2.1 – *Architecture générale d'un système de dialogue.*

Les modules nécessaires pour construire un système de dialogue interactif sont les suivants :

– **Reconnaissance de la parole (Automatic Speech Recognition, ASR) :**
Ce module a pour but de transformer le signal audio de l'entrée de l'utilisateur vers une séquence transcrite de mots. Les systèmes utilisés récemment en dialogue sont basés sur des approches probabilistes.
Ces approches consistent à rechercher la séquence de mots la plus vraisemblablement prononcée étant donné le signal de parole émis par le locuteur. Le module de reconnaissance probabiliste est souvent constitué de deux sous-modules : l'un représente le modèle acoustique et l'autre le modèle de langage.
Plusieurs systèmes de reconnaissance de la parole peuvent être cités : CUED-HTK (Woodl et al., 1998), CMU Sphinx (Lee et al., 1992), Microsoft Whisper (Huang et al., 1995). On peut citer également des systèmes de reconnaissance français notamment le système du LIMSI (Gauvain et al., 1994), SPEERAL par le LIA (Nocera et al., 2002), ou encore ANTS (Fohr et al., 2004) par le LORIA.

Bien que ces systèmes soient assez performants, leurs capacités dans le cadre d'un système de dialogue dépend de plusieurs paramètres spécifiques à la tâche. Des erreurs de transcription peuvent être causées par le bruit provenant d'une conversation téléphonique, des phrases mal formées ou incomplètes (parole spontanée), etc.

– **Compréhension de la parole (Spoken Langage Understanding, SLU) :**
Dans un système de dialogue interactif, le module de compréhension de la parole est intermédiaire entre la reconnaissance de la parole et le gestionnaire de dialogue. Son rôle est de "traduire" le texte transcrit (à la sortie de l'ASR) vers des séquences de concepts sémantiques. Autrement dit, le but du processus de compréhension est d'extraire une liste d'hypothèses d'étiquettes de concepts à partir d'une phrase en entrée. Ces concepts représentent la sémantique de l'information existant dans la phrase en entrée.
Les techniques de développement de ce module ont évolué ces dernières années puisqu'on est passé de méthodes à base de règles à des méthodes statistiques. Ce module est le cœur de notre étude, donc il sera étudié plus en détail dans la section suivante.

– **Gestionnaire de dialogue (Dialogue Act Manager, DAM) :**
Le gestionnaire de dialogue est l'unité centrale dans un système de dialogue, il est pour le système ce qu'un processeur est pour un ordinateur. Son rôle est de prendre les décisions sur la procédure à suivre par le système. Il définit les futures actions à suivre au cours du dialogue en se basant sur les flux d'informations qu'il a récupéré dans l'état actuel du dialogue et en se connectant avec des sources extérieures (base de données, web...). La performance du DAM dépend de sa capacité à prendre la meilleure décision parmi plusieurs choix possibles (et de la finesse de ces choix).
Plusieurs approches peuvent être utilisées pour l'implémentation d'un gestionnaire de dialogue tel que les approches orientées plans (eg. le système TRAINS (Allen et al., 1994)), le modèle d'agenda (eg. le système OLYMPUS (Bohus et al., 2007)). D'autres approches reposent sur le principe d'état d'information (eg. le système DIPPER (Bos et al., 2003)) ou sur des POMDP (eg. le système BUDS de l'université de Cambridge (Thomson and Young, 2010) ou l'approche proposée au LIA (Pinault and Lefèvre, 2011)).

– **Génération de texte (Natural Language Generation, NLG) :**
Pour pouvoir communiquer avec l'utilisateur et continuer le déroulement d'un dialogue, les actions décidées par le module DAM doivent être partagées avec l'utilisateur. Dans un premier temps le générateur de texte est chargé de générer un texte qui contient la réponse ou l'information que le module DAM veut communiquer avec l'utilisateur. Généralement ce module est basé sur des patrons textuels prédéfinis ou en utilisant des méthodes plus complexes de génération d'énoncés (Walker et al., 2007).

– **Synthèse de parole (Text-to-Speech Synthesis, TTS) :**
Le message généré par le module NLG est transmis à l'utilisateur en utilisant un système de génération de la parole. Cela peut être basé sur des patrons pré-enregistrés ou en utilisant un système de synthèse vocale.

En général, ces modules sont construits à part, puis mis en série pour construire le système de dialogue complet. Sachant que la plupart des modules utilisés récemment sont des modèles statistiques capables de générer une liste d'hypothèses possibles, la performance du système de dialogue peut être améliorée en considérant des listes de n-meilleures transcriptions ou annotations. Les canaux de transmission entre les différents modules peuvent devenir plus riches en information. Les sorties de l'ASR sont remplacées par des listes de n-meilleures transcriptions ou par des treillis de mots. De la même manière les sorties du SLU peuvent être présentées comme des listes de n-meilleures hypothèses à l'entrée du DAM. Cette architecture dite en "fat pipeline" permet une optimisation globale sur les différents composants du système de dialogue (Williams, 2008).

Dans le cadre de cette thèse, nous nous concentrons sur l'étude du système de compréhension de la parole que nous allons présenter plus en détail. La suite de ce chapitre est organisée de la manière suivante : une introduction aux systèmes de compréhension de la parole est présentée dans la section 2.2 et les notions de représentation sémantique pour la compréhension sont définies dans la section 2.3. Les méthodes utilisées pour obtenir un système de compréhension sont présentées dans la section 2.4 tout en distinguant les deux types d'approches utilisées pour les construire ; les approches linguistiques dans la section 2.4.1 et les approches issues de l'apprentissage automatique dans la section 2.4.2. Pour finir nous présentons les mesures qui peuvent être appliquées pour évaluer et comparer des systèmes de compréhension dans la section 2.5.

2.2 Le système de compréhension de la parole

Comme son nom l'indique, le rôle du système de compréhension de la parole est d'extraire le sens d'un signal de parole. Ce sens est représenté par des étiquettes sémantiques (concepts) qui décrivent la sémantique d'un domaine. Deux choix doivent être faits lors de la construction d'un module de compréhension. Le premier concerne la manière dont le sens est représenté (la représentation sémantique). Cette représentation peut être sous plusieurs formes telles que des requêtes SQL, les arbres sémantiques ou le langage des concepts. Le second concerne la méthode utilisée pour obtenir la représentation sémantique à partir d'un message vocal transcrit. Ces méthodes peuvent être classées dans deux grandes familles : les approches basées sur les connaissances linguistiques (Allen, 1987) et les approches issues de l'apprentissage automatique (Pieraccini et al., 1991; Minker et al., 1996).

2.3 Représentations sémantiques pour la compréhension

Dans un système de dialogue, le sens d'un message vocal est l'ensemble des informations utiles pour réaliser une tâche spécifique. Ainsi, dans le cas d'un système de recherche d'information, le sens d'un énoncé utilisateur peut être représenté directement sous la forme qui permet d'interroger la base de données pour avoir l'information souhaitée. Dans la campagne d'évaluation des systèmes de compréhension ATIS (Air Travel Information System) (Zue et al., 1992), les sens des énoncés utilisateurs ont été représentés par des requêtes SQL. De la même manière, le système d'AT&T (How May I Help You ?) (Gorin et al., 1997) représente le sens d'un message vocal simplement par la destination (le service) vers laquelle l'utilisateur doit être redirigé (call routing).

Cette représentation du sens a l'avantage d'être liée directement à l'application qu'on souhaite faire du message dans le système de dialogue. Cependant, cette représentation manque de généricité car elle est très liée au système pour lequel elle a été produite et donc un nouveau système de compréhension nécessite un nouveau système d'annotation. De plus, cette représentation rend l'évaluation du système de compréhension difficile car elle nécessite un système de dialogue complet pour évaluer la compréhension.

Pour dépasser les limites de cette représentation simple, d'autres formalismes ont été proposés pour représenter le sens d'un message. Les réseaux sémantiques proposés par Woods (Woods, 1970) sont utilisés pour la représentation sémantique. Dans ce formalisme, la sémantique d'une phrase prend la forme d'entités (les nœuds du réseau) et de relations typées entre ces entités. Plusieurs formalismes ont été proposés pour représenter les réseaux sémantiques, tels que KL-ONE (Brachman, 1979), PropBank (Kingsbury and Palmer, 2003) et FrameNet (Baker et al., 1998).

Pour illuster l'approche par frames sémantiques de FrameNet, l'exemple de la frame *Activity* est donné ci-dessous :

```
Activity

Definition:
This is an abstract frame for durative activities, in which the
Agent enters an ongoing state of the Activity, remains in this
state for some Duration of Time, and leaves this state either
by finishing or by stopping.

Semantic Type: Non-Lexical Frame

FEs:

Core:
Agent [Agent]
Semantic Type: Sentient The Agent is engaged in the Activity.
```

```
Core Unexpressed:
Activity [Act] This FE identifies the Activity in which the
Agent is engaged.

Non-Core:
Duration [Dur]
Semantic Type: Duration This FE identifies the amount of Time
an Activity takes.

Manner [Manr]
Semantic Type: Manner Any description of the activity which
is not covered by more specific FEs, including secondary effects

Place [Place]
Semantic Type: Locative_relation This FE identifies the Place where
the Activity occurs.

Time [Time]
Semantic Type: Time This FE identifies the Time when the Activity occurs.

Frame-frame Relations:

Inherits from: Process
Is Inherited by: Apply_heat
Perspective on:
Is Perspectivized in:
Uses:
Is Used by:
Subframe of:
Has Subframe(s): Activity_abandoned_state, Activity_done_state,
                 Activity_finish, Activity_ongoing,
                 Activity_paused, Activity_prepare,
                 Activity_ready_state, Activity_resume,
                 Activity_start, Activity_stop
Precedes:
Is Preceded by:
Is Inchoative of:
Is Causative of:
See also:
```

Cette frame modélise un scénario où une personne (Agent) réalise une activité (Activity) pendant une période de temps. Cette représentation a l'inconvénient de ne pas pouvoir tenir compte du contexte de dialogue et donc ne fournit pas toujours une représentation suffisamment précise du sens d'un message.

Corpus	Langue	Domaine	Phrases	Concepts
MEDIA	français	réservation touristique	18k	99
TELDIR	allemand	horaires de train	22k	23
LUNA	polonais	information de transport	12k	200
LUNA	italien	support technique	5k	40
PlanResto	français	réservation de restaurant	12k	59
ATIS	anglais	réservations de billets d'avion	6k	17

TABLE 2.1 – *Une comparaison entre des corpus de dialogue pour des domaines et des langues différents.*

Les limites des réseaux sémantiques ont encouragé un passage vers une représentation plus spécifique du sens qui repose sur une annotation précise sur l'ensemble de données disponibles pour une tâche spécifique. Cette annotation est basée sur un ensemble de concepts qui représente la sémantique du domaine.

Plusieurs corpus de dialogue (avec une annotation conceptuelle) sont disponibles dans différentes langues et pour plusieurs tâches. La taille des corpus de dialogue en général est de l'ordre de quelques dizaines de milliers d'énoncés utilisateur. La complexité de ces corpus est caractérisée par la tâche concernée et par le nombre de concepts sémantiques utilisés pour étiqueter le corpus. Dans le tableau 2.1, nous comparons quelques corpus de dialogue disponibles pour des tâches différentes : MEDIA (Bonneau-Maynard et al., 2005), TELDIR (Aust et al., 1995), LUNA (Aust et al., 1995), PlanResto (Sadek et al., 1996) et ATIS (Hemphill et al., 1990).

L'annotation sémantique pour le corpus MEDIA présente l'originalité d'être segmentale. Chaque phrase est découpée en plusieurs segments, chaque segment correspond à une unité sémantique (un concept). Pour certaines approches une annotation au niveau du mot est nécessaire pour apprendre le modèle. Une manière de transformer l'annotation au niveau des segments vers une annotation au niveau des mots est d'utiliser le formalisme BIO (Begin Inside Outside) (Ramshaw and Marcus, 1995).

Ce formalisme consiste à étiqueter le premier mot de chaque segment conceptuel par l'étiquette "B-concept", les autres mots du segment par l'étiquette "I-concept" et les mots qui n'appartiennent à aucun segment par "O" ou "NULL". Ce formalisme permet à la fois d'obtenir une annotation mot à mot des phrases et de marquer les frontières entre les segments conceptuels successifs. Par exemple la phrase "from Phoenix to San Diego April first" du corpus 'ATIS sera annotée en utilisant le format BIO comme :

```
from B_departure_city
Phoenix I_departure_city
to B_arrival_cit
San I_arrival_city
Diego I_arrival_city
April B_departure_date_month
first B_departure_date_day_number
```

Les corpus annotés ont l'avantage de donner une bonne représentation sémantique du sens contenu dans la phrase mais sont coûteux en temps et en expertise humaine. Le coût de tels corpus est lié à la collecte, la transcription et l'annotation de données. En général ces corpus sont collectés pour une langue et un domaine précis et donc leur généralisation n'est pas toujours évidente.

Plusieurs outils ont été utilisés pour aider l'annotation humaine des corpus de dialogue. Par exemple l'outil Semantizer développé au LIMSI (Bonneau-Maynard and Rosset, 2003) est une interface d'annotation qui permet d'annoter les messages de dialogue par plusieurs annotations (concept, mode, valeur, specifieur). Cet outil a été utilisé lors de l'annotation du corpus MEDIA (Bonneau-Maynard et al., 2005). Ce corpus est le corpus sur lequel sont basées toutes les expériences présentées dans cette thèse. Il sera décrit en détail dans la section 5.2.1.

La tâche d'annotation peut être une tâche assez complexe surtout lorsqu'il s'agit d'une annotation de l'ordre de centaine de concepts différents. La difficulté et le coût lié à l'annotation sont les motivations de ce travail de thèse qui cherche à porter un corpus existant dans une langue donnée vers une nouvelle langue afin de minimiser le coût lié à la création d'un nouveau corpus.

2.4 Approches pour la compréhension de la parole

Dans un système de dialogue la compréhension automatique de la parole consiste à transformer la transcription d'un message vocal vers une forme compatible avec le gestionnaire de dialogue. Pour réaliser cette opération, deux types d'approches ont été proposés : les approches linguistiques et les approches issues de l'apprentissage automatique.

2.4.1 Approches linguistiques

Les approches linguistiques pour la compréhension de la parole sont fondées sur une analyse syntaxique et/ou sémantique de la phrase à étiqueter. En général cette approche associe à chaque mot tous ses sens possibles et ensuite garde l'hypothèse qui donne un sens cohérent pour la totalité de la phrase. Cette approche, telle que décrite par (Allen, 1987) permet de construire l'arbre sémantique associé à une phrase donnée en utilisant la logique du domaine. La FIGURE 2.2 présente un exemple de l'arbre sémantique de la phrase "je voudrais réserver un hôtel à Paris le 5 juin".

Ce formalisme est basé sur un ensemble de catégories caractérisées par une fonction et un argument. Dans l'exemple précédent, la catégorie "ville" a la fonction "à" et l'argument "Paris".

Chomsky (Chomsky, 1957, 1959) a été un des premiers à tenter de représenter la langue de cette manière en utilisant des grammaires dites "formelles" (formal grammars). Ces grammaires permettent d'analyser une langue donnée en utilisant un nom-

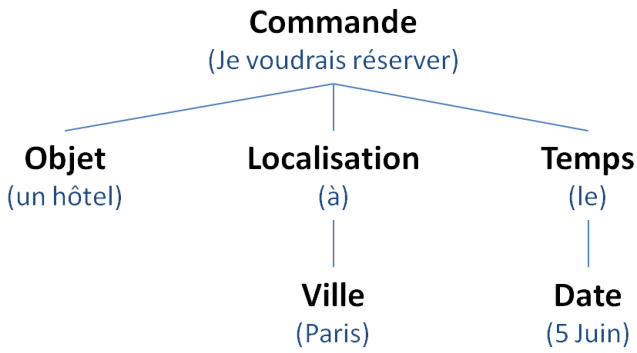

FIGURE 2.2 – *Arbre sémantique associé à la phrase "je voudrais réserver un hôtel à Paris le 5 Juin".*

bre fini de règles qui représentent les différentes associations possibles des mots de cette langue. Les grammaires formelles sont classées par Chomsky et Schützenberger (Chomsky and Schützenberger, 1963) selon leur expressivité en 4 types : les grammaires non restreintes (unrestricted grammars), les grammaires contextuelles (context-sensitive grammars), les grammaires algébriques hors contexte (context-free grammars) et enfin les grammaires régulières (regular grammars). Les grammaires non restreintes n'imposent aucune contrainte alors que les autres types de grammaires sont de plus en plus restrictives.

Malgré le fait que les grammaires hors contexte soient les plus utilisées dans les approches linguistiques du traitement automatique des langues, ces grammaires restent incapables de modéliser le langage naturel d'une manière fine et correcte. C'est pourquoi Woods (Woods, 1970) a proposé d'utiliser des grammaires à base de réseaux de transitions augmentées (Augmented Transition Network Grammars, ATNG) pour mieux modéliser le langage naturel. Ces grammaires représentent une combinaison entre les connaissances sémantiques sensibles au contexte et les informations syntaxiques.

Dans le contexte de dialogues nous avons affaire à de la parole et donc à des phrases spontanées qui ont leurs spécificités (répétition, hésitation, reprise, ...) et donc ce sont souvent des phrases agrammaticales. Pour faire face à ce genre de phrase, des grammaires basées sur les aspects sémantiques ont été proposées telles que les grammaires de cas (Case Grammar) aussi appelées "cadres sémantiques" (Fillmore, 1985). Ces grammaires sont basées sur un ensemble de cas qui représentent les relations entre un verbe et ses composants nominaux. Selon (Bruce, 1975), un cas est une relation entre un verbe et un de ses arguments. L'ensemble de cas qui couvre une langue donnée peut être nommée "une grammaire de cas".

Dans le cadre de la compréhension de la parole, les approches par grammaire de cas peuvent être utilisées pour fournir un support sémantique lors de l'analyse de ces phrases. Le sens d'un énoncé est déterminé par une analyse de cas qui détermine le sens de la requête. Plusieurs travaux ont appliqué les grammaires de cas dans le cadre de systèmes de dialogue (Matrouf et al., 1989; Lamel et al., 1999; Bennacef et al., 1996; Villaneau et al., 2004). Le tableau 2.2 donne un exemple de grammaire de cas pour la

CASEFRAME : flight-time
KEYWORDS : vol, voyager, aller, partir
from : (quitte, de)@city
to : (à, pour, vers)@city
torelative-departure-time : (partir+)avant, après
departure-time : (partir+)@hour-minute
CASEFRAME : @city
{city : dallas, boston, atlanta, ...}
CASEFRAME : @hour-minute
{...}

TABLE 2.2 – *Exemple de cadre sémantique pour la tâche ATIS.*

tâche ATIS en français (Bennacef et al., 1994).

Dans le tableau 2.2 nous observons trois cadres sémantiques. Le premier (flight-time) est associé à plusieurs éléments (from, to, torelative-departure-time, departure-time). Chaque élément est associé à un ou plusieurs attributs. Ces attributs ont des valeurs définies (avant, après) ou sont associés à un autre cadre (@city).

D'autres modèles de représentation de l'information sont les réseaux sémantiques proposés par (Quillian, 1968). Un réseau sémantique est un graphe dont les sommets représentent des concepts sémantiques et les arcs représentent les relations entre ces concepts.

Les grammaires formelles ont évolué vers des grammaires stochastiques pour prendre en compte l'ambigüité d'analyse liée aux spécificités de la parole. Une grammaire hors contexte probabiliste peut donc estimer la probabilité d'une analyse en se basant sur un corpus d'apprentissage. Un exemple de ces grammaires est l'analyseur linguistique TINA développé à l'institut de technologie du Massachussetts (MIT) (Seneff, 1989). Cet analyseur utilise une grammaire hors contexte transformée de façon automatique en un automate portant des probabilités sur les arcs, permettant d'avantager les constructions les plus courantes.

Les approches linguistiques pour la compréhension de la parole sont limitées par la structure des messages de l'utilisateur. Ces messages sont souvent agrammaticaux ou inachevés et donc une partie importante de l'information contenue dans ces messages est perdue dans l'analyse linguistique. En plus, les systèmes de reconnaissance de la parole génèrent un nombre important d'erreurs sur les messages de parole ce qui encourage le passage aux approches stochastiques qui peuvent être plus robustes aux erreurs de transcription et s'adaptent mieux aux spécificités de l'oral.

2.4.2 Approches issues de l'apprentissage automatique

Les approches issues de l'apprentissage automatique constituent une alternative efficace aux méthodes à base de règles et de grammaire car elles réduisent la nécessité de

l'expertise humaine : un ensemble de phrases sémantiquement annotées est suffisant pour l'apprentissage des modèles de compréhension. Plusieurs approches statistiques ont été proposées pour la compréhension de la parole (Schwartz et al., 1996; Wang and Acero, 2006; Lefèvre, 2007; Hahn et al., 2008; Suendermann et al., 2009a; Hahn et al., 2010). Un autre atout majeur de ces approches est la disponibilité d'un ensemble d'hypothèses évaluées (matérialisée par des treillis, des listes n-meilleures, etc) qui élargissent le passage de l'information allant du système de compréhension vers le gestionnaire de dialogue. Le gestionnaire de dialogue peut ensuite ajuster sa décision en tenant compte de certaines ambiguïtés possibles (si par exemple deux hypothèses ont des scores proches) et permet de composer des hypothèses scorées pour les états de dialogue sur plusieurs sorties de compréhension tour par tour.

2.4.2.1 Les modèles Markoviens

Les modèles Markoviens peuvent être utilisés pour apprendre un étiqueteur sémantique. Nous pouvons distinguer deux catégories de ces modèles : les modèles génératifs (les réseaux bayesiens dynamiques, DBN dont les modèles de Markov cachés, HMM sont un cas particulier) et les modèles discriminants (les modèles de Markov à maximum d'entropie, MEMM (McCallum et al., 2000) et les champs aléatoires conditionnelles, CRFs (Lafferty et al., 2001)).

Une comparaison graphique entre ces modèles est donnée dans la FIGURE 2.3. Cette figure montre que les HMM et les MEMM sont représentés par des graphes orientés. En revanche le graphe des CRFs n'est pas orienté, ce qui veut dire que les probabilités conditionnelles sur les concepts dépendent de l'ensemble des observations lors de l'utilisation des CRFs et non uniquement de l'observation courante.

Ainsi un avantage déterminant des CRFs (discriminants) par rapport aux HMM (génératifs) est la possibilité d'utiliser l'ensemble des observations d'une séquence pour prédire une étiquette. Donc, pour attribuer une étiquette à une observation, non seulement l'historique immédiat d'observations est pris en compte mais toutes les observations précédentes et suivantes.

Les modèles de Markov cachés (Hidden Markov Model, HMM) sont des modèles génératifs qui peuvent être appliqués à une tâche de compréhension sous la forme :

- un espace de mots W
- un ensemble de concepts C
- une loi qui définit la probabilité du concept initial $P(c_1)$
- un ensemble de probabilités de transition d'un concept C vers un concept suivant C' : $P(C'|C)$
- un ensemble de probabilités d'un mot selon les concepts : $P(W|C)$

Dans les modèles HMM, la probabilité d'un concept à l'instant t ne dépend que du concept à l'instant précédent et un mot produit à l'instant t ne dépend que du concept à ce même instant. Considérant tous ces critères, la probabilité d'une séquence de mots peut être calculée comme :

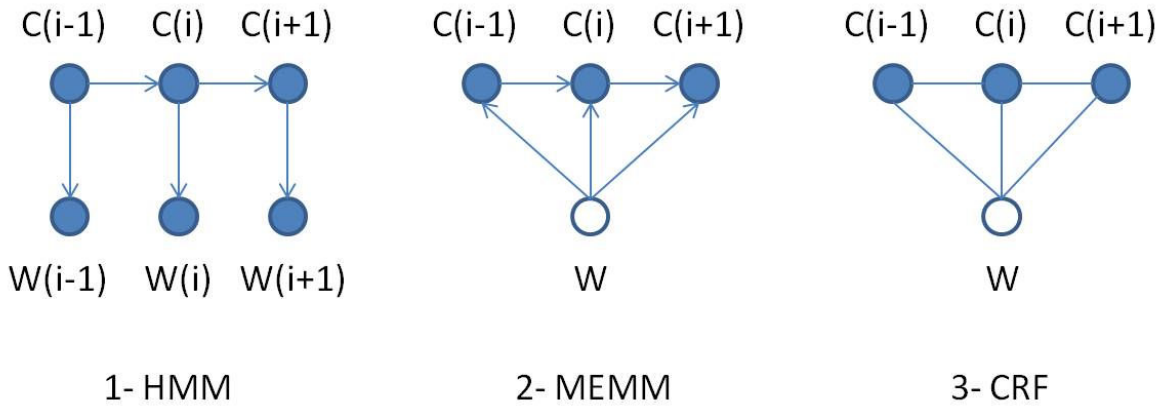

FIGURE 2.3 – *Représentation graphique des modèles HMM, MEMM et CRFs.*

$$\hat{C} = \text{argmax}_c P(C \mid W)$$

l'application de la règle de Bayes sur cette représentation donne :

$$\hat{C} = \text{argmax}_c \frac{P(W \mid C).P(C)}{P(W)}$$

W est constante tout au long du décodage donc sa probabilité n'influencera pas le calcul de la fonction argmax_c et donc l'équation peut être reformulée comme :

$$P(C|W) = \text{argmax}_{c_i}[P(w_1|c_1)P(c_1)\prod_{t=1}^{T} P(w_t|c_t)P(c_t|c_{t-1})]$$

Une alternative aux HMM est l'utilisation des modèles discriminants. Nous pouvons distinguer deux modèles Markoviens discriminants qui ne diffèrent qu'au niveau de la normalisation de la probabilité totale qu'ils représentent. Le premier est le modèle de Markov à maximum d'entropie (Maximum Entropy Markov Model, MEMM) (McCallum et al., 2000) pour lequel la normalisation est appliquée au niveau du mot, alors que le second modèle repose sur les champs aléatoires conditionnels (Conditional Random Fields, CRFs) (Lafferty et al., 2001) pour lesquels la normalisation s'applique au niveau de la phrase.

Dans les deux cas, la représentation de ces modèles est décrite comme une probabilité conditionnelle d'une séquence de concept $C = c_1, ..., c_N$ étant donnée une séquence de mots $W = w_1, ..., w_N$. Cette probabilité peut être calculée comme suit :

$$P(C|W) = \frac{1}{Z}\prod_{n=1}^{N} H(c_{n-1}, c_n, w_{n-2}^{n+2})$$

avec

$$H(c_{n-1}, c_n, w_{n-2}^{n+2}) = \sum_{m=1}^{M} \lambda_m \cdot h_m(c_{n-1}, c_n, w_{n-2}^{n+2})$$

H est un modèle log-linéaire fondé sur des fonctions caractéristiques $h_m(c_{n-1}, c_n, w_{n-2}^{n+2})$ qui représentent l'information extraite du corpus d'apprentissage. Les poids λ du modèle log-linéaire sont estimés lors de l'apprentissage et Z est un terme de normalisation défini par les modèles MEMM tel que :

$$Z = \prod_{n=1}^{N} \sum_{\hat{c}_n} H(c_{n-1}, c_n, w_{n-2}^{n+2})$$

Pour les modèles CRFs la normalisation est au niveau de la phrase donc Z est défini telle que :

$$Z = \sum_{\hat{c}_1^N} \prod_{n=1}^{N} H(c_{n-1}, c_n, w_{n-2}^{n+2})$$

Les modèles markoviens ont été largement utilisés pour la tâche de la compréhension de la parole. Le système CHRONUS (Conceptual Hidden Representation Of Natural Unconstrained Speech) (Pieraccini et al., 1991) est un des premiers systèmes de compréhension par une approche statistique. Le système utilise un modèle HMM dont les états représentent les concepts. Les séquences de mots associées à un concept donné sont également modélisées par un processus markovien représenté par un modèle de langage n-grammes conditionné par le concept. CHRONUS est évalué sur la tâche ATIS (Air Travel Information System) décrite en détail dans (De Mori, 1997). Plusieurs systèmes de compréhension à base d'HMM ont été proposés par le LIMSI comme le modèle proposé par (Minker et al., 1996) pour la tâche ATIS et celui proposé par (Maynard and Lefèvre, 2002) pour une tâche de réservation téléphonique de billets de train pour le système de dialogue ARISE (Lamel et al., 2000).

Les CRFs ont le plus souvent été utilisés dans le domaine du traitement automatique des langues, pour étiqueter des séquences d'unités linguistiques. Ces modèles possèdent à la fois les avantages des modèles génératifs et des modèles discriminants de l'état de l'art. En effet, comme les classifieurs discriminants, ils peuvent manipuler un grand nombre de descripteurs, et comme les modèles génératifs, ils intègrent des dépendances entre les étiquettes de sortie et prennent une décision globale sur la séquence. Plusieurs travaux ont proposé d'utiliser les CRFs pour modéliser un système de compréhension de la parole (Wang and Acero, 2006; Lefèvre, 2007; Raymond and Riccardi, 2007; Hahn et al., 2009). Par ailleurs, plusieurs outils sont disponibles pour apprendre ces modèles tels que CRF++ (Kudo, 2005) et Wapiti (Lavergne et al., 2010).

2.4.2.2 Approche par traduction automatique (SMT)

La compréhension de la parole peut être vue comme la traduction d'une phrase en langage naturel vers une autre phrase en langage de concept. Cette interprétation transforme la tâche de compréhension en une tâche de traduction et donc les approches de traduction automatique peuvent être utilisées pour réaliser cette tâche.

Pour cela (Macherey et al., 2009; Hahn et al., 2010) ont proposé d'utiliser une méthode de traduction statistique standard qui regroupe plusieurs modèles : un modèle de traduction automatique statistique à base de segments (Phrase-Based Statistical Machine Translation, PB-SMT), un modèle de langage et un modèle de réordonnancement.

Plusieurs outils sont disponibles pour apprendre un tel système (MOSES (Koehn et al., 2007) pour le modèle de traduction et SRILM (Stolcke, 2002) pour le modèle de traduction). Vu que la traduction automatique représente un des autres aspects important de cette thèse, nous consacrons un chapitre complet à état de l'art de cette méthode.

2.4.2.3 Les transducteurs à états finis (FST)

Les transducteurs à états finis (Finite State Transducers, FST) sont des automates à états finis avec des sorties (émissions de symboles). Le transducteur à états finis prend un mot en entrée et le transforme en un autre mot en sortie contrairement à l'automate à état fini (accepteur) qui simplement accepte ou rejette le mot. En plus de l'entrée et de la sortie, des poids peuvent être associés à chaque transaction (Weighted Finite State Transducer, WFST). Cette caractéristique rend les FSTs utilisables pour plusieurs tâches de traitement automatique de la langue (Roche and Schabes, 1995; Mohri, 1997; Mohri et al., 2002; Yvon et al., 2004).

(Raymond et al., 2006) ont proposé d'utiliser les FSTs pour une tâche de compréhension de la parole suivant la proposition initiale de (Pereira and Wright, 1997). Dans cette approche le processus de transformation d'une séquence de mots w_1^N vers une séquence de concepts c_1^N est obtenu par un automate à état fini. Cet automate est la combinaison de trois automates :

- le premier $\lambda_{w_1^N}$ représente la phrase à étiqueter w_1^N,
- le deuxième λ_{w2c} représente les transductions de mots vers des concepts et
- le troisième λ_{LM} représente un modèle de langage de concepts.

Le meilleur étiquetage est le meilleur chemin λ_{SLU} tel que :

$$\lambda_{SLU} = \lambda_{w_1^N} \circ \lambda_{w2c} \circ \lambda_{LM}$$

Cette approche a l'avantage de pouvoir utiliser l'ensemble des opérations définies sur les FST, notamment les opérations de compositions, de déterminisation et de recherche de plus court chemin.

Ces opérations permettront de combiner le processus de compréhension au processus de reconnaissance ou à une tâche de traduction comme cela sera montré dans le chapitre 6.

Un modèle de compréhension peut être obtenu facilement par cette approche grâce à la disponibilité d'outils libres pour construire les FSTs tels que OpenFst (Allauzen et al., 2007).

2.4.2.4 Les machines à vecteur de support (SVM)

En plus des méthodes d'étiquetage séquentiel probabilistes il est aussi possible de recourir aux approches de classification automatiques. Ces approches considèrent le problème de la compréhension d'une séquence comme un problème de classification de l'ensemble des mots de cette séquence.

Ces méthodes cherchent à trouver la fonction qui minimise les erreurs de classification. La méthode la plus utilisée récemment pour cette tâche repose sur les SVMs (Vapnik, 1982).

Les machines à vecteurs de support (Support Vector Machine, SVM) proposées par Vapnik (Vapnik, 1982; Cherkassky, 1997) sont des classifieurs à valeurs réelles utilisés dans de nombreuses tâches d'apprentissage automatique. Ces classifieurs découpent le problème de classification en deux sous-problèmes : la transformation non-linéaire des entrées et le choix d'une séparation linéaire optimale.

Le premier problème vient du besoin de se projeter dans un espace où les données peuvent être séparées linéairement. Pour ce faire, une transformation basée sur un noyau est appliquée pour projeter les données dans un espace de grande dimension (voir la FIGURE 2.4). Le noyau est une fonction (linéaire, polynomiale ou gaussienne) qui retourne la valeur du produit scalaire entre deux vecteurs d'entrée $K(X, Y) = \Phi(X).\Phi(Y)$.

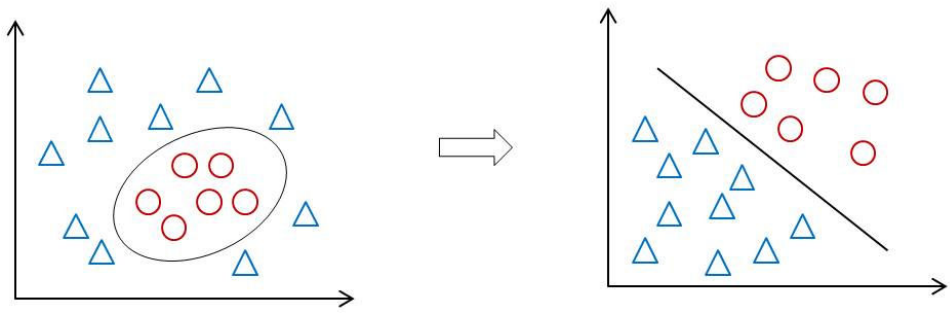

FIGURE 2.4 – *Projection des données dans un espace de grande dimension.*

Le deuxième problème est de trouver la marge maximale, qui représente la distance entre la frontière de séparation et les points les plus proches. Ces derniers sont appelés vecteurs de support, dans le sens qu'ils supportent la limite séparatrice.

Sachant que la plupart des problèmes de classification sont linéairement séparables et que les méthodes SVM ont l'avantage de pouvoir traiter un grand nombre de données et d'assez grande dimension, elles sont utilisées pour des tâches différentes telles que la catégorisation de texte (Joachims, 1998), la reconnaissance du locuteur (Ho and Moreno, 2004) ou encore l'identification du texte dans les images (Herv et al., 2001).

Les SVMs peuvent être utilisées avec succès pour une tâche de compréhension de la parole. YAMCHA est un exemple de système de compréhension fondé sur les SVM (Kudoh and Matsumoto, 2000). Ce système consiste à entrainer un classifieur pour chaque paire de classes conceptuelles à reconnaitre.

Vu que les SVM sont des classifieurs binaires, donc dans un environnement de K classes, $K(K-1)/2$ classifieurs binaires sont construits pour prendre en compte toutes les paires de classes. L'étiquetage final est un vote pondéré sur les différents classifieurs binaires. Une approche comparable mais permettant un décodage conceptuel à partir d'une représentation structurée a été proposé par l'université de Cambridge (Mairesse et al., 2009). Des outils libres sont disponibles pour apprendre ces modèles (eg. l'outil YAMCHA (Kudo and Matsumoto, 2001)).

Plusieurs travaux (Schapire and Singer, 2000; Haffner et al., 2003) ont montré que les méthodes de classification peuvent être un moyen efficace pour extraire des concepts à partir d'une phrase transcrite. Cette approche présente l'avantage d'être robuste au bruit de transcription généré par le système de reconnaissance et aux effets dus à la parole spontanée.

D'autre part, cette approche réduit la contribution humaine dans la création du modèle du fait que le corpus d'apprentissage doit contenir, pour chaque phrase, l'ensemble des concepts qu'il implique sans avoir besoin de définir les mots sur lesquels ces concepts reposent. On peut noter que, à part les CRFs, cette approche donne la meilleure performance comparée aux approches stochastiques présentées précédemment (Hahn et al., 2010).

2.5 Evaluation des systèmes de compréhension de la parole

L'évaluation des systèmes de compréhension permet de comparer des systèmes entre eux et aussi d'optimiser les paramètres choisis pour un modèle donné. On peut distinguer deux types d'évaluation : l'évaluation de séquences d'étiquettes et l'évaluation globale du message.

Dans le premier type chaque mot ou séquence de mots du corpus de test est associé à une étiquette sémantique et l'évaluation consiste à aligner cet étiquetage avec les étiquettes produites par le modèle évalué afin de calculer le nombre d'insertion de substitution et d'omission de concepts.

Pour cela le score CER (Concept Error Rate) peut être utilisé. Le CER est l'équivalent du taux d'erreur en mots (Word Error Rate, WER) mais au niveau des concepts. Le WER est dérivé de la distance de Levenshtein (Levenshtein, 1965) en appliquant la

comparaison au niveau des mots au lieu des caractères. Donc le CER peut être défini comme le ratio de la somme des concepts omis, insérés et substitués sur le nombre de concepts dans la référence. Le CER représente un pourcentage d'erreurs, donc il est d'autant meilleur qu'il est petit.

$$CER = \frac{count(Ins) + count(Del) + count(Sub)}{count(concepts\ de\ la\ référence)} * 100$$

D'autres mesures peuvent être utilisées pour une évaluation globale du message dans les cas où on ne considère pas la notion séquentielle. Dans ces cas les mesures de précision et de rappel sont utilisées pour évaluer les systèmes. La précision représente le pourcentage de concepts corrects trouvés par le système sur la totalité des concepts générés par le système.

$$précision = \frac{count(concepts\ corrects\ trouvés)}{count(concepts\ trouvés)} * 100$$

Le rappel représente le pourcentage de concepts corrects retrouvés parmi les concepts attendus dans la référence.

$$rappel = \frac{count(concepts\ corrects\ trouvés)}{count(concepts\ à\ trouver)} * 100$$

Enfin, la F-mesure représente une mesure unique qui permet de combiner à la fois précision et rappel. L'efficacité globale du système selon la F-mesure peut être définie par l'équation suivante :

$$F - mesure = \frac{2 * précision * rappel}{précision + rappel}$$

2.6 Conclusion

Dans ce chapitre nous avons introduit les systèmes de dialogue homme-machine de manière générale et nous avons présenté plus particulièrement le module de compréhension de la parole, objet d'étude principal de cette thèse. Le développement de ce module peut être réalisé par des approches linguistiques ou des approches issues de l'apprentissage automatique.

Ces dernières ont montré des bonnes performances pour la tâche de compréhension. Ces approches minimisent le besoin en expertise humaine nécessaire pour développer les modèles linguistiques et nécessitent uniquement un corpus d'apprentissage constitué d'énoncés annotés. Le choix d'une approche ou d'une autre dépend énormément de la taille des corpus disponibles et aussi de la complexité de la représentation sémantique.

Plusieurs travaux ont comparé les performances des différentes approches statistiques pour des tâches similaires (Hahn et al., 2008, 2010). Les conclusions obtenues par ces travaux montrent que l'approche à base de CRFs est (jusqu'à présent) la plus performante pour une tâche d'étiquetage séquentiel. C'est pourquoi nous avons fait le choix d'utiliser cette méthode dans les expériences réalisées pour cette thèse.

Chapitre 3

La traduction automatique

Sommaire

3.1 Introduction 38
3.2 Architectures des systèmes de traduction 39
3.2.1 Architectures linguistiques 39
3.2.2 Architectures computationnelles 40
3.3 La traduction automatique probabiliste 41
3.3.1 Modèle de langage 42
3.3.2 Modèle de traduction 44
3.3.2.1 Traduction à base de mots 45
3.3.2.2 Traduction à base de segments 48
3.3.3 Modèle log-linéaire 49
3.3.4 Décodage 51
3.3.5 Approche hiérarchique pour la traduction automatique 53
3.4 Outils pour la traduction automatique probabiliste 55
3.5 Evaluation des systèmes de traduction 56
3.6 Conclusion 57

3.1 Introduction

La traduction automatique est un domaine de la linguistique computationnelle qui consiste à traduire un texte (écrit ou oral) depuis une langue source vers une langue cible. Un logiciel de traduction automatique analyse le texte dans la langue source (texte à traduire) et génère automatiquement le texte correspondant dans la langue cible (texte traduit) à l'aide d'un ordinateur.

Dans les années 50, la recherche en traduction automatique portait sur la traduction littérale, à savoir la traduction mot à mot, sans prise en compte des règles linguistiques. Le système démontré à l'Université de Georgetown (connu sous le nom de l' "Expérience de Georgetown IBM") en 1950 représente la première tentative systématique visant à créer un système de traduction automatique utilisable.

Des recherches sont également menées en Europe et aux États-Unis, tout au long des années 50 et au début des années 60.

En 1966, aux États-Unis, le rapport ALPAC (Automatic Language Processing Advisory Committee) fait une estimation prématurément négative de la valeur des systèmes de traduction automatique, et des perspectives offertes par ceux-ci, mettant fin au financement et à l'expérimentation dans ce domaine pour la décennie suivante.

C'est seulement à la fin des années 70 que des tentatives sérieuses sont à nouveau entreprises, parallèlement aux progrès de l'informatique et des technologies des langues.

Cette période a vu aussi le développement de systèmes de transfert et l'émergence des premières tentatives commerciales. Des sociétés comme Systran et Metal sont persuadées que la traduction automatique est un marché viable et utile. Elles mettent sur pied des produits et services de traduction automatique reliés à un serveur central. Mais les problèmes sont nombreux : des coûts élevés de développement, une lexicographie demandant un énorme travail, des difficultés pour proposer de nouvelles combinaisons de langues, et l'inaccessibilité de tels systèmes pour l'utilisateur moyen.

Dans les années 80, beaucoup de travaux sur la représentation morphologique, syntaxique et sémantique sont réalisés au Japon. En 1991, le premier modèle de traduction automatique statistique est proposé par IBM ; la traduction repose sur des modèles numériques appris à partir de nombreuses phrases alignées source et cibles.

Aujourd'hui la recherche en traduction automatique statistique est extrêmement vaste et devient de plus en plus populaire parmi les chercheurs en traitement automatique de la langue. Un historique de la traduction automatique ainsi qu'une vue globale sur ses différents paradigmes peuvent être trouvés dans (Dorr et al., 1999) et (Hutchins, 2007).

Ce chapitre présente un bref état de l'art de la traduction automatique. En premier lieu, nous présentons les différentes architectures des systèmes de traduction tout en distinguant deux catégories d'architectures : architectures linguistiques 3.2.1 et architectures computationnelles 3.2.2. Dans le cadre de cette thèse nous nous intéressons à la

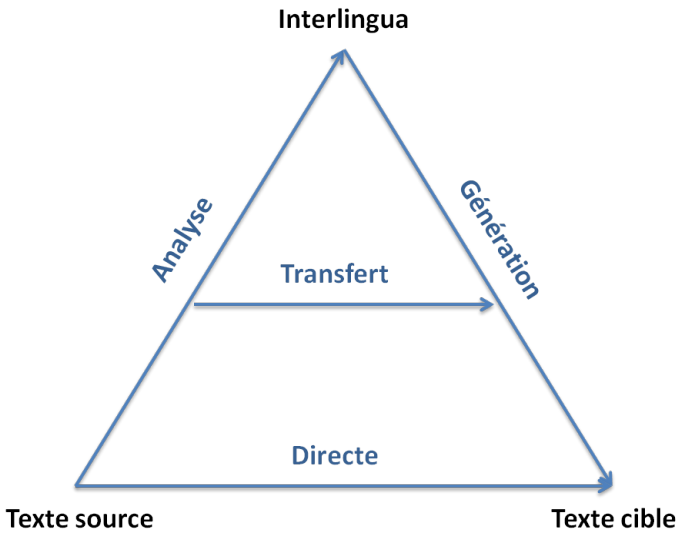

FIGURE 3.1 – *Le triangle de Vauquois.*

traduction automatique probabiliste que nous présentons en détail dans la section 3.3. Enfin avant de conclure ce chapitre nous présentons des outils utilisés pour la traduction automatique probabiliste dans la section 3.4 et les métriques utilisées pour évaluer la traduction automatique dans la section 3.5.

3.2 Architectures des systèmes de traduction

Dans (Boitet, 2008) l'auteur différencie deux classes d'architecture du système de traduction automatique : architecture linguistique et architecture computationnelle. L'architecture linguistique d'un système de traduction est caractérisée par la représentation qu'utilise ce système durant le processus de traduction, tandis que l'architecture computationnelle d'un système de traduction est caractérisée par les méthodes de calcul lors du processus de traduction.

3.2.1 Architectures linguistiques

Les différentes architectures linguistiques possibles d'un système de traduction ont été représentées par le fameux triangle de Vauquois (Vauquois, 1968) (Figure 3.1). Ces architectures peuvent être définies de la manière suivante :

– **Système de traduction directe :**
Dans ces systèmes la traduction d'une langue vers une autre est réalisée en une seule étape sans passer par une analyse de la phrase source ni par un processus de génération de la phrase cible. Cette traduction est effectuée en utilisant une

structure plus ou moins évoluée d'un dictionnaire bilingue qui traduit un mot ou une séquence de mots d'une langue vers une autre (Hutchins and Somers, 1992). Les systèmes de traduction directe ont plusieurs limites ; le manque d'analyse syntaxique (à plat ou hiérarchique) ou lexicale étant le principal problème.

Certains défauts de ces systèmes peuvent être comblés par les approches de traduction "indirecte" dans laquelle une analyse linguistique des phrases sources est une étape indispensable du processus de traduction. Les systèmes de traduction de ce type peuvent être classés dans deux catégories : les systèmes de transfert et les systèmes de traduction par interlingua.

- **Système de transfert :**
  Dans ces systèmes la traduction se fait en trois étapes : une étape d'analyse de la phrase source, une étape de transfert et une étape de génération de la phrase cible. Ces systèmes nécessitent des connaissances linguistiques dans la langue source et la langue cible (représentées par des dictionnaires, des grammaires, etc.) et en plus des règles de transformation qui relient la langue source avec la langue cible (une table qui contient les règles de transfert entre les deux langues) (Arnold et al., 1993).

- **Système de traduction par interlingua :**
  Cette approche implique l'utilisation d'une représentation intermédiaire (ou langue pivot, l'interlingua) pour passer d'une langue vers une autre. Le processus de traduction d'une langue source vers une langue cible consiste à analyser les phrases de la langue source pour les transférer vers la langue intermédiaire, puis ensuite générer des phrases traduites dans la langue cible à partir de cette traduction pivot.

  Cette approche a l'avantage d'être facilement applicable pour des traductions multilingues (ie au-delà d'une paire de langues unique). Pour pouvoir traduire entre n langues différentes nous avons besoin de n analyseurs et n générateurs au lieu de $n(n-1)$ systèmes de transfert. Cependant les analyseurs et les générateurs pour les systèmes de traduction par interlingua sont "théoriquement" plus complexes et plus difficiles à établir.

3.2.2 Architectures computationnelles

L'architecture computationnelle d'un système de traduction est définie par la méthode utilisée pour réaliser la traduction. En général, les systèmes de traduction peuvent être regroupés en trois catégories d'architecture computationnelle :

- **Traduction à base de règles :**
  Pour construire un système de traduction à base de règles, nous demandons à des spécialistes d'établir des règles qui définissent la traduction d'une langue à une autre et ensuite ces connaissances sont encodées dans le système. Ces connais-

sances peuvent être sous plusieurs formes comme des dictionnaires bilingues ou des règles de transfert syntaxiques d'une langue à une autre. Plusieurs types de grammaires peuvent être utilisés dans ces méthodes pour modéliser les langues comme les LFG (Lexical Functional Grammars) (Kaplan and Bresnan, 1995), les HPSG (Head-driven Phrase Structure Grammar) (Pollard, 1985) ou les TAG (Tree Adjoining Grammar) (Kroch and Joshi, 1985).

– **Traduction à base d'exemples :**
Ces systèmes (que l'on peut classer, au même titre que les systèmes statistiques, dans la catégorie des systèmes empiriques ou numériques), considèrent que le modèle de traduction est construit automatiquement à partir d'un bitexte (texte traduit en deux langues) assez grand sans avoir besoin d'expertise pour établir des règles. Le bitexte (mémorisé par le système) constitue un ensemble d'exemples à partir desquels le système construit une hypothèse cible à partir d'un texte source en concatenant des morceaux d'exemples issus de la mémoire initiale (Nagao, 1984; Langlais et al., 2008).

– **Traduction statistique (probabiliste) :**
Les systèmes de traduction statistiques ont besoin aussi d'un bitexte pour être conçus. Ce bitexte permet d'estimer une fonction de densité calculant la probabilité qu'un texte cible soit la traduction d'un texte source. Lors de la traduction de nouveaux textes, les systèmes de traduction statistiques cherchent à maximiser cette probabilité pour trouver la traduction la plus probable.
Les données bilingues sont de plus en plus disponibles avec l'utilisation croissante de sites web bilingues et les traductions multilingues de certains documents officiels comme les débats parlementaires européens par exemple. Grâce à la disponibilité de ces données, la traduction statistique a connu une croissance importante au cours des dernières années. C'est l'approche utilisée par plusieurs systèmes de traduction en ligne comme le système de Google [1] et celui de Microsoft [2].

3.3 La traduction automatique probabiliste

Le processus de traduction consiste à transformer une phrase dans la langue source S, vers une phrase dans la langue cible T. Sachant que chaque phrase est composée d'une séquence de mots, ces phrases peuvent être définies comme : $s = s_1, ..., s_i, ...s_I$ et $t = t_1, ..., t_j, ...t_J$ où s_i représente le mot dans la position i de la phrase source et t_j représente le mot dans la position j de la phrase cible.

Brown et al (Brown et al., 1990, 1993) ont fait l'hypothèse que chaque phrase d'une langue peut être la traduction de chaque phrase d'une autre langue avec des probabilités différentes. Donc à chaque paire de phrases (s, t) est attribuée une probabilité

1. http ://translate.google.com
2. http ://www.bing.com/translator

$P(t \mid s)$, cette probabilité correspond à la probabilité que t soit la traduction de s. La traduction probabiliste cherche à trouver la traduction la plus correcte (la plus probable) pour une phrase s en appliquant des règles de décision avec une perte minimale (Och, 2005) :

$$\hat{t} = \text{argmax}_t P(t \mid s)$$

L'application de la règle de Bayes sur cette représentation donne :

$$\hat{t} = \text{argmax}_t \frac{P(s \mid t).P(t)}{P(s)}$$

Vu que le texte source et le même tout au long du décodage, sa probabilité $P(s)$ n'influencera pas le calcul de la fonction argmax_c, donc l'équation peut être reformulée comme :

$$\hat{t} = \text{argmax}_t P(s \mid t).P(t)$$

Cette représentation peut être décomposée en deux modèles séparés : le modèle de traduction $P(s \mid t)$ et le modèle de langage $P(t)$. Le modèle de traduction (Translation Model, TM) cherche à trouver la meilleure traduction d'une phrase en entrée, tandis que le modèle de langage (Language Model, LM) attribue des probabilités aux hypothèses de traduction selon la probabilité de l'occurrence de cette hypothèse dans la langue cible. Le processus de décodage (traduction) consiste à trouver la meilleure traduction en considérant les probabilités données par le modèle de langage sur toutes les hypothèses de traduction.

3.3.1 Modèle de langage

Les modèles statistiques de langage ont été utilisés dans plusieurs applications du traitement automatique du langage naturel (Natural Language Processing, NLP), tel que la reconnaissance de la parole (Chen and Goodman, 1996; Roark et al., 2007), l'étiquetage des catégories syntaxiques (Part-Of-Specch, POS), la recherche d'informations (Information Retrieval, IR) (Song, Fei et al., 1999), etc.

Le modèle de langage tente d'estimer la probabilité d'une séquence de mots dans une langue cible. Il représente la probabilité qu'un mot ou une séquence de mots apparaisse dans une langue. En traduction automatique la plupart des modèles de langage sont des modèles n-grammes. Pour une phrase e un modèle de langage est défini comme la probabilité jointe des mots $w_1, w_2, ...w_n$ qui la compose :

$$P(e) = P(w_1 w_2 ... w_n)$$

En appliquant la règle de probabilité des chaînes, cette représentation peut être décomposée comme :

$$P(e) = P(w_1)P(w_2 \mid w_1)P(w_3 \mid w_1 w_2)...P(w_n \mid w_1...w_{n-1})$$

En pratique, ce modèle est simplifié en faisant l'hypothèse qu'un mot w_i dépend uniquement des $n-1$ mots précédents. Ainsi nous obtenons un modèle de langage bigrammes avec $n = 2$ et un modèle trigrammes avec $n = 3$. Un modèle de langage trigrammes par exemple peut être représenté comme :

$$P(e) = P(w_1)P(w_2 \mid w_1)\prod_{i=3}^{n} P(w_i \mid w_{i-1}w_{i-2})$$

L'estimation des probabilités du modèle de langage statistique se fait par rapport à un corpus d'apprentissage monolingue. Si ce corpus est d'une très grande taille on peut considérer qu'il représente la langue en général.

La probabilité d'apparition d'un mot est généralement estimée en utilisant le critère de maximum de vraisemblance (Maximum Likelihood Estimation, MLE). Pour un modèle trigrammes, la probabilité d'un mot w_3 de la séquence $w_1w_2w_3$ est donnée comme :

$$P(w_3 \mid w_1w_2) = \frac{count(w_1w_2w_3)}{count(w_1w_2)}$$

où $count(w_1w_2w_3)$ est le nombre d'occurences de la suite $w_1w_2w_3$ dans le corpus d'apprentissage, et $count(w_1w_2)$ est le nombre d'occurences de la suite w_1w_2 dans le corpus d'apprentissage.

Cette modélisation pose un problème lorsqu'on rencontre des mots ou des n-grammes non observés dans les données d'apprentissage. Quelle que soit la taille du corpus d'apprentissage, il y a toujours des mots absents dans le corpus (mots hors vocabulaire). Même si tous les mots d'un n-grammes apparaissent dans le corpus d'apprentissage, il est possible que leur apparition ne soit pas du même ordre recontré au moment de la traduction et donc le n-grammes sera inconnu. Pour ces cas, l'estimation de probabilité de base est de 0 et donc par conséquent, une probabilité nulle sera attribuée à toute séquence de mots contenant un mot hors vocabulaire et à tous les n-grammes inconnus.

Une manière de résoudre ce problème est d'utiliser un algorithme de lissage (smoothing algorithm). Plusieurs algorithmes de lissage ont été proposés afin de généraliser les modèles de langage (Chen and Goodman, 1996). Le lissage de Laplace (add one) consiste à ajouter la fréquence 1 à tous les n-grammes. Le lissage par repli (backoff) (Katz, 1987) consiste à utiliser un modèle du même ordre (trigrammes, bigrammes,...) lorsqu'un n-gramme apparaît dans le corpus d'apprentissage, et utiliser

un modèle d'ordre inférieur (unigramme par exemple) dans le cas contraire. Ce processus peut être itéré jusqu'au zéro-gramme, qui consiste à attribuer une constante indépendante du mot. Le lissage par interpolation part du même principe que le lissage par repli, sauf qu'il consiste à combiner un modèle avec des modèles d'ordre inférieur systématiquement même lorsque la fréquence des mots n'est pas 0.

Perplexité

La performance d'un modèle de langage varie selon plusieurs critères : la taille du corpus sur lequel il a été appris, la longueur des n-grammes et l'algorithme de lissage utilisé. La perplexité (Bahl et al., 1977) est la mesure la plus couramment employée pour évaluer les modèles de langages. Elle est le facteur de l'entropie $H(LM)$ d'un modèle LM utilisé dans la prédiction des mots rencontrés. Elle peut être obtenue de la manière suivante :

$$PP(LM) = 2^{H(LM)}$$

Dans le cadre de l'utilisation des modèles de langage, le facteur d'entropie peut être calculé comme suit :

$$H(LM) = -\frac{1}{n}\log P(w_1 w_2 ... w_n)$$

Cette valeur est mesurée sur un texte non vu au cours de l'apprentissage (test-set perplexity). Elle est différente de la perplexité du modèle sur ses données d'apprentissage (training-set perplexity).

La perplexité peut être considérée comme le facteur de branchement moyen du langage cible. Un modèle de langage ayant une perplexité X est équivalent, en terme de perplexité, à un langage qui comporterait X mots équiprobables.

3.3.2 Modèle de traduction

L'alignement de mots est une étape importante pour apprendre un système de traduction statistique. En général, pour deux phrases parallèles où l'une est la traduction de l'autre, l'alignement de mots peut être défini comme une correspondance entre les mots de ces deux phrases.

Les modèles de traduction peuvent être regroupés en deux types, selon l'unité de base de la traduction (de l'alignement) : les modèles à base de mots et les modèles à base de segments.

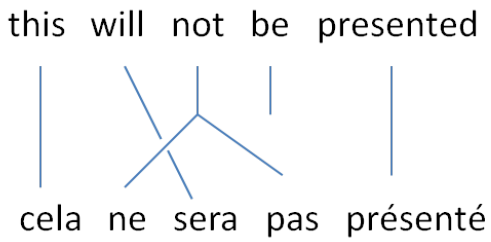

FIGURE 3.2 – *Exemple d'un alignement de mots.*

3.3.2.1 Traduction à base de mots

Les tout premiers systèmes de traduction statistique sont basés sur les mots comme unité élémentaire. La traduction entre deux phrases peut être vue comme une correspondance "alignement" entre les mots de la phrase source et ceux de la phrase cible. Les mots de la phrase source peuvent être réordonnés selon l'ordre d'occurrence des mots cible auxquels ils sont alignés.

Les modèles IBM (Brown et al., 1993) sont les premiers modèles probabilistes proposés pour créer des modèles de traduction à base de mots. (Brown et al., 1990) considère que chaque mot de la phrase cible est aligné uniquement à un seul mot de la phrase source.

Ils font aussi l'hypothèse que chaque mot de la phrase source peut correspondre à un certain nombre de mots de la phrase cible.

Ce nombre est nommé la fertilité du mot. Dans l'exemple présenté dans la Figure 3.2, le mot "not" possède une fertilité de 2. D'autre part le mot "be" n'est aligné à aucun mot et donc a une fertilité de 0. Le mot "NULL" peut être utilisé pour être aligné avec les mots dont la fertilité est 0.

Pour estimer le modèle de traduction $P(s \mid t)$ nous devons représenter les alignements entre les mots des deux langues. En général le nombre de mots entre les deux phrases (source et cible) n'est pas identique, donc une variable a qui représente le nombre d'alignements possibles est utilisée, ainsi que le modèle de traduction qui peut être défini comme :

$$P(s \mid t) = \sum_{a} P(s, a \mid t)$$

$P(s, a \mid t)$ peut être défini comme :

$$P(s, a \mid t) = P(I \mid t) \sum_{i=1}^{I} P(a_i \mid s_1^{i-1}, a_1^{i-1}, I, t).P(s_i \mid a_1^{i}, s_1^{i-1}, I, t)$$

sachant que :

- $I =$ la longueur de la phrase source s
- $s_i =$ le mot de la phrase s dans la position i
- $a_i =$ la position du mot cible aligné à s_i

IBM a défini cinq modèles différents pour estimer $P(s \mid t)$ chaque modèle est basé sur les paramètres estimés par le modèle précédent. Ces modèles peuvent être définis comme suit :

- **IBM-1** est un simple modèle de traduction à base de mots. Il suppose que tous les réordonnancements sont équiprobables. Il représente uniquement un modèle lexical.

$$P_{IBM-1}(a_i \mid s_1^{i-1}, a_1^{i-1}, J, t_1^m) = \frac{1}{m+1}$$

  La longueur de la phrase cible m a été incrémentée de 1 pour prendre en compte le mot vide.

- **IBM-2** suppose que la position d'un mot est déterminée par la position du mot auquel il est aligné et par la longueur des phrases sources et cibles.

$$P_{IBM-2}(a_i \mid s_1^{i-1}, a_1^{i-1}, J, t_1^m) = P(a_i \mid i, J, m)$$

  Un modèle HMM similaire au modèle IBM-2 a été proposé par (Vogel et al., 1996). Ce modèle modélise la distance entre les alignements plutôt que de modéliser la position des mots. (Och and Ney, 2003) ont montré que les modèles HMM sont légèrement meilleurs que IBM-2.

$$P_{HMM}(a_i \mid s_1^{i-1}, a_1^{i-1}, J, t) = P(a_i - a_i)$$

  Un exemple d'alignement produit par les modèles IBM-1, IBM-2, et HMM est donné dans la figure 3.3. Dans cette figure chaque mot source est aligné à un mot cible par un "anneau" d'alignement. Le nombre d'anneau d'alignement d'une phrase correspond au nombre d'alignements possibles.

- **IBM-3** introduit —en plus de la traduction lexicale et des règles d'alignement— des choix de fertilité pour chaque mot de la phrase cible. Ce modèle permet aussi d'aligner des mots cibles avec le mot source "NULL".

- **IBM-4** prend en compte l'ordre des mots dans chaque phrase. Toutefois les modèles IBM 1 à 4 sont déficients : ils assignent une probabilité non nulle à des réordonnancements impossibles et ne prennent pas en compte les positions cibles déjà couvertes.

- **IBM-5** est une version améliorée du modèle IBM-4 et donne des résultats plus satisfaisants. Le modèle IBM-5 est considéré comme un modèle de référence dans la littérature.

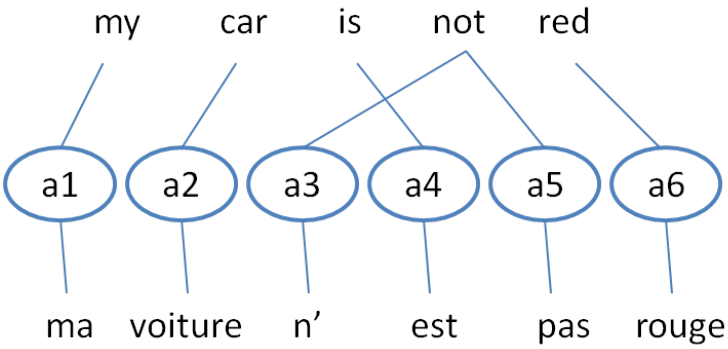

FIGURE 3.3 – *Exemple d'alignement produit par les modèle IBM-1, IBM-2 et HMM*

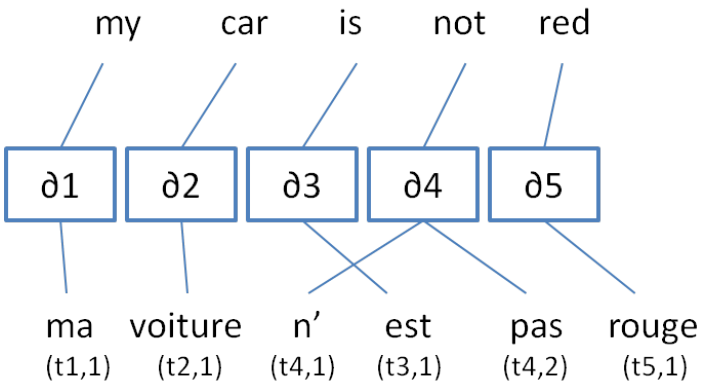

FIGURE 3.4 – *Exemple d'alignement produit par les modèle IBM-3, IBM-4 et IBM-5.*

Un exemple d'alignement produit par les modèles IBM-3, IBM-4, et IBM-5 est donné dans la Figure 3.4. Dans les modèles IBM 3 à 5 les alignements sont présentés par des tablettes qui représentent un alignement du type 1-à-n.

Ces trois modèles regroupent les alignements associés au même mot cible dans une seule "tablette". Par exemple le mot "not" est aligné à une seule tablette au lieux d'être aligné à deux anneaux d'alignement.

Plus d'informations sur les modèles IBM sont données dans (Och, 2003b), et une comparaison systématique entre ces modèles est donnée dans (Och and Ney, 2003).

Les modèles IBM permettent d'obtenir un alignement dans les deux directions de traduction séparément. (Liang et al., 2006) ont proposé d'aligner les deux directions en même temps en favorisant les accords entre les deux modèles "Alignment by agreement".

Des modèles discriminants peuvent aussi être utilisés pour obtenir des alignements en mots (Moore et al., 2006; Allauzen and Wisniewski, 2010). Ces modèles facilitent l'intégration d'informations arbitraires dans l'alignement, tels que des connaissances

sur le domaine ou des propriétés grammaticales.

3.3.2.2 Traduction à base de segments

Dans la traduction à base de mot, les choix de traduction sont pris sur chaque mot séparément. Cette traduction est loin d'être parfaite surtout dans les cas où un mot de la langue source est traduit par plusieurs mots de la langue cible.

D'un point de vue linguistique, la traduction d'un mot d'une langue vers une autre ne dépend pas uniquement de ce mot mais aussi des mots qui l'entourent dans la même phrase. En plus de cela, la traduction à base de mot nécessite une segmentation des phrases en mots ce qui n'est pas évident dans certaines langues comme l'arabe ou le chinois.

La traduction à base de segments est une solution efficace pour faire face à ce problème. Dans cette approche les segments sont les unités de base de la traduction à la place des mots. Une séquence de mots de la phrase source est la traduction d'une séquence de mots de la phrase cible. Chaque séquence de la phrase source forme un segment (phrase en anglais).

Cette approche est largement utilisée aujourd'hui. Sa performance par rapport à l'approche de traduction à base de mots vient de sa capacité de gérer mieux le réordonnancement de mot et de mieux traduire les expressions idiomatiques. Lorsque le modèle de traduction est appris sur une grande quantité de données, les segments peuvent couvrir des phrases entières et donc obtenir des traductions parfaites pour ces phrases.

Plusieurs travaux on été dédiés à la traduction probabiliste à base de segments (Phrase-Based Statistical Machine Translation, PB-SMT) (Koehn et al., 2003; Marcu and Wong, 2002; Och and Ney, 2004). Dans cette approche la notion du mot NULL n'existe pas et donc chaque segment non vide est traduit par un segment non vide dans la langue cible.

Le processus de traduction d'une phrase source vers une phrase cible se fait généralement en trois étapes :

1. La phrase source est découpée en segments.
2. Chaque segment source est traduit vers la langue cible, séparément, en fonction de sa probabilité de traduction $\Phi(\hat{s}_i \mid \hat{t}_i)$.
3. Les segments traduits sont permutés selon l'ordre naturel de la langue cible. La phrase traduite est obtenue avec une probabilité de distorsion $d(a_i - b_{i-1})$, sachant que a_i et b_{i-1} représente le début et la fin du ième segment source traduit en ième segment cible.

Le modèle de traduction dans cette approche représente à la fois la probabilité de traduire le segment et le modèle de distorsion. Ce modèle peut être défini pour une phrase de I segments comme :

$$P(s \mid t) = \prod_{i=1}^{I} \Phi(\hat{s}_i \mid \hat{t}_i).d(a_i - b_{i-1} - 1)$$

La probabilité de traduction d'un segment $\Phi(\hat{s}_i \mid \hat{t}_i)$ est estimée en utilisant le critère de maximum de vraisemblance (Maximum Likelihood Estimation, MLE) et peut être définie comme :

$$\Phi(\hat{s}_i \mid \hat{t}_i) = \frac{count(t_i, s_i)}{\sum count(t_i, s_i)}$$

Afin de pouvoir estimer la probabilité de traduction d'un segment, les alignements en segments doivent être disponibles. L'alignement en segment se fait selon plusieurs méthodes, mais nous décrivons dans la suite une approche implémentée dans un système état de l'art (MOSES, décrit dans la section 3.4), qui se fait à partir de deux corpus parallèles en deux étapes :

- La symétrisation des alignements en mots des phrases dans les deux sens de traduction : un alignement symétrisé en mots peut être obtenu en combinant deux alignements en mots. Cette combinaison peut être réalisée de plusieurs manières (union, intersection ou en utilisant des méthodes heuristiques). La figure 3.5 présente un exemple d'un alignement symétrisé obtenu par l'union de deux alignements en mots.
- L'extraction de segments : l'extraction de correspondance (extraction de segments) peut être réalisée, à partir d'un alignement symétrique, en respectant un critère principal. Dans un segment, chaque séquence de mots source est aligné mutuellement à une séquence de mots cible et aucun mot de ces deux séquences n'est aligné à d'autres mots qui n'appartiennent pas au segment. Plus de détails sur ces critères sont donnés dans (Och et al., 1999; Zens et al., 2002). La figure 3.6 donne un exemple d'un segment consistant et d'un autre inconsistant.

3.3.3 Modèle log-linéaire

Malgré le fait que le modèle de traduction à base de segment tel que présenté jusque là soit plus performant que le système à base de mots, cette performance peut être augmentée en étendant ce modèle vers un modèle log-linéaire.

Le modèle de traduction à base de segment contient trois modèles :
- le modèle de traduction $\Phi(\hat{s} \mid \hat{t})$,
- le modèle de langage $P(t)$ et
- le modèle de réordonnancement ou de distorsion d.

La meilleure traduction e_{best} générée par cette méthode peut être obtenue par :

$$e_{best} = \text{argmax}_t \prod_{i=1}^{I} \Phi(\hat{s}_i \mid \hat{t}_i).d(a_i - b_{i-1} - 1) \prod_{i=1}^{t} P(t_i \mid t_1 ... t_{i_1})$$

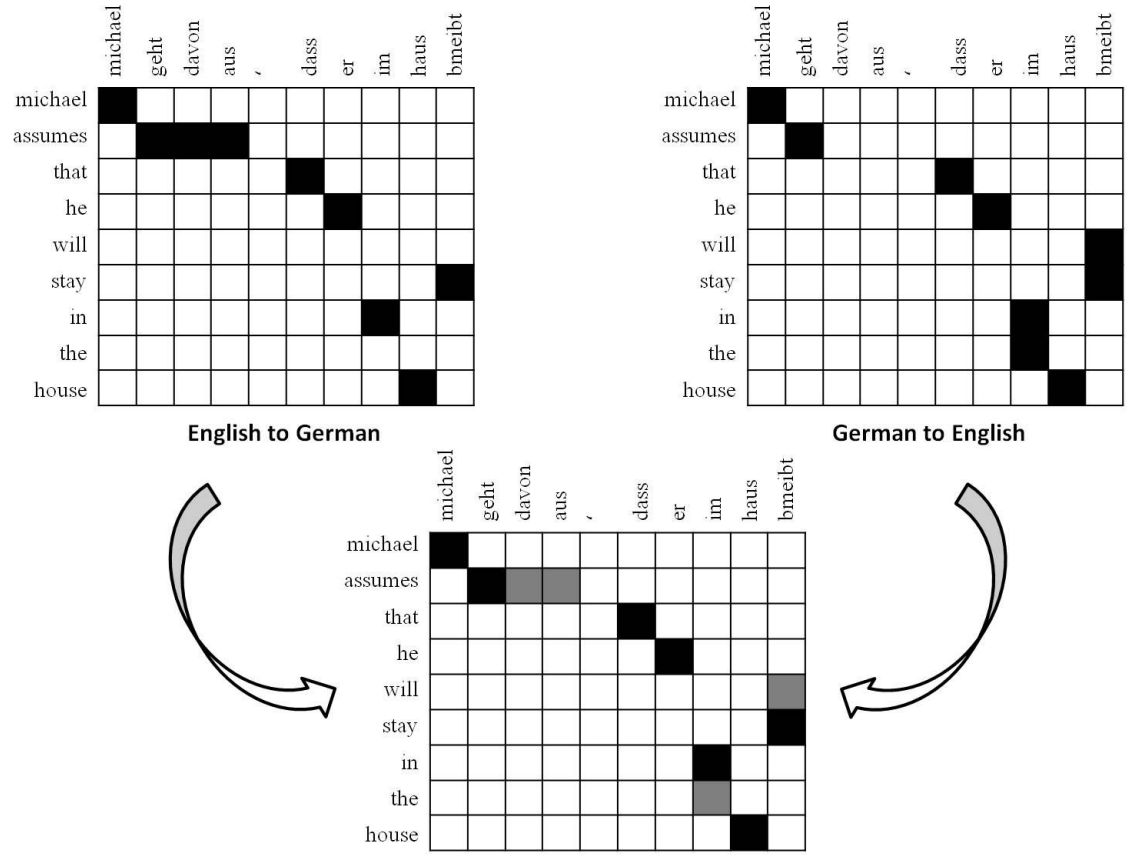

FIGURE 3.5 – *Un alignement symétrisé obtenu par l'union d'alignements en mots.*

Ce modèle peut nous conduire vers des situations où les mots de sortie correspondent bien aux mots d'entrée mais la phrase au final n'est pas très compréhensible. Dans des cas similaires nous souhaitons donner au modèle de langage un poids plus important. Des poids différents peuvent être donnés à chaque modèle (λ_Φ pour le modèle de traduction, λ_d pour le modèle de distorsion et λ_{LM} pour me modèle de langage). Ainsi la meilleure traduction peut être définie comme :

$$e_{best} = \text{argmax}_t \prod_{i=1}^{I} \Phi(\hat{s}_i \mid \hat{t}_i)^{\lambda_\Phi} . d(a_i - b_{i-1} - 1)^{\lambda_d} \prod_{i=1}^{t} P(t_i \mid t_1 ... t_{i_1})^{\lambda_{LM}}$$

L'ajout des poids dans ce modèle part de considérations pratiques et non pas de rigueur mathématique. Cependant, un modèle log-linéaire (utilisée dans différentes approches de traduction automatique) peut être utilisé et décrit comme suit :

$$P(x) = \exp \sum_{i=1}^{n} \lambda_i h_i(x)$$

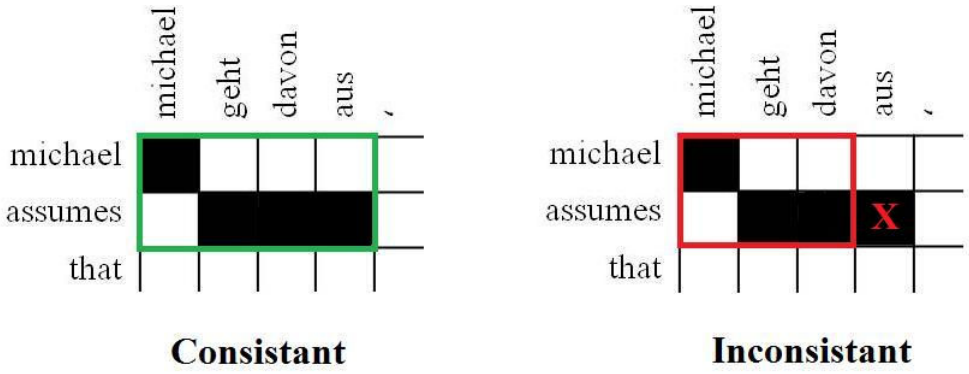

FIGURE 3.6 – *Exemple d'extraction de segments*

L'équation du calcul de la meilleure traduction peut être adaptée à cette formule en considérant :

- nombre de fonction $n = 3$;
- la variable aléatoire $x = (s, t, a_i, b_i)$;
- la fonction $h_1 = \log \Phi$;
- la fonction $h_2 = \log d$;
- la fonction $h_3 = \log P_{LM}$;

Cette équation peut être présentée plus clairement sous la forme :

$$P(t, a \mid s) = \exp(\lambda_\Phi \sum_{i=1}^{I} \log \Phi(\hat{s}_i \mid \hat{t}_i) + \lambda_d \sum_{i=1}^{I} \log d(a_i - b_{i-1} - 1) + \lambda_{LM} \sum_{i=1}^{t} \log P(t_i \mid t_1 ... t_{i_1})$$

En plus d'avoir l'avantage d'attribuer des poids aux différents composants, le modèle log-linéaire permet d'étendre le modèle avec des composants supplémentaires.

Plusieurs algorithmes ont été proposés pour régler les poids du modèle log-linéaire afin d'obtenir une traduction optimale. L'apprentissage à taux d'erreur minimum (Minimum Error Rate Training, MERT) (Och, 2003a) est traditionnellement utilisé pour optimiser la traduction (notamment le score BLEU) selon l'algorithme présenté dans le tableau 3.1.

3.3.4 Décodage

Un problème assez important dans la traduction automatique est le décodage (la traduction). Après avoir estimé les modèles de traduction et de langage, nous cherchons à traduire une phrase de la langue source vers la langue cible.

Le décodage consiste à trouver la meilleure traduction en considérant ces modèles. Ce problème a été défini comme un problème NP-complet (Knight, 1999; Zaslavskiy

Algorithme de MERT

Input : initial weights λ_1^L
  repeat
    Generate N-best repository with current λ_1^L
    for all dimensions l **do**
      for all sentences s **do**
        Compute upper enveloppe and error statistics
      end for
      Merge error statistics
      Search for optimal γ and determine error reduction Δ_{el}
      $\acute{\lambda}_l \leftarrow \lambda_l + \gamma$ where γ is the change of weight for a given dimension
    end for
    $\lambda_l \leftarrow \acute{\lambda}_{\widehat{l}}$ **with** $\widehat{l} = argmin_l(\Delta_{el})$
  until convergence
  return λ_1^L

TABLE 3.1 – *Le pseudo code de l'algorithme MERT pour l'optimisation du modèle log-linéaire.*

et al., 2009). La complexité vient du fait qu'il existe plusieurs choix de traduction pour une phrase source donnée. Un très grand nombre de probabilités doit être estimé afin de pouvoir trouver la meilleure traduction.

Pour minimiser le coût de génération de traductions, plusieurs traductions partielles peuvent être développées en parallèle et seule la meilleure est complétée. Un grand nombre d'approches de traduction probabiliste utilisent l'algorithme de recherche par piles "stack decoder" pour le décodage. Dans cet algorithme les phrases cibles sont construites de gauche à droite et les traductions partielles couvrant le même nombre de mots sources sont regroupées par pile.

L'algorithme de recherche "A*" est un algorithme de recherche par piles utilisé par (Och et al., 2001) pour le décodage en traduction automatique. Dans cet algorithme chaque hypothèse partielle est associée à un coût défini par la somme d'un score préfixe Q et d'un score heuristique H. Le score préfixe Q représente le coût de génération de cette hypothèse partielle et le score heuristique H représente le coût nécessaire pour obtenir l'hypothèse totale de la traduction à partir de cette hypothèse partielle. Ce score est généralement obtenu par le calcul du produit du coût de génération de tous les mots source non couverts.

En premier la pile est initialisée par une hypothèse vide. Ensuite la pile est triée par coût croissant et l'hypothèse en tête de la pile est étendue en couvrant une position source supplémentaire. Après chaque extension, les nouvelles hypothèses partielles générées sont incorporées à la pile avec les coûts accompagnés, et la pile est triée à nouveau. Le décodeur continue l'extension jusqu'à ce que tous les mots sources soient couverts. Une fois que tous les mots source sont couverts, le décodeur renvoie l'hypothèse qui est en tête de la pile comme solution.

```
Beam Search Algorithm
place empty hypothesis into stack 0
for all stacks 0 ... n − 1 do
    for all hypotheses in stack do
        for all translation options do
            if applicable then
                create new hypothesis
                place in stack
                recombine with existing hypothesis if possible
                prune stack if too big
            end if
        end for
    end for
end for
```

TABLE 3.2 – *Le pseudo code de l'algorithme de recherche en faisceau.*

Le coût de cette hypothèse est le coût de génération Q (le score H est de zéro car tous les mots source sont couverts). Cette hypothèse représente l'hypothèse à coût minimal et elle est considérée comme la meilleure traduction.

L'algorithme de "A*" permet de trouver de bonnes solutions mais il est très coûteux en temps lorsqu'il s'agit de phrases longues. (Tillmann and Ney, 2003) ont proposé d'utiliser l'algorithme de recherche en faisceau (beam search) qui permet d'optimiser le processus de recherche en réduisant le besoin de mémoire.

Cet algorithme n'étend qu'un nombre limité d'hypothèses qui sont les hypothèses les plus prometteuses et donc élague les hypothèses non pertinentes. Le pseudo-code de cet algorithme est donnée dans le tableau 3.2 et l'algorithme est illustré dans la figure 3.7.

Deux méthodes d'élagage sont couramment utilisées :

- élagage par seuil : les hypothèses dont la probabilité est inférieure à n fois la probabilité de la meilleure hypothèse dans la même pile sont élaguées.
- élagage par histogrammes : uniquement les n meilleures hypothèses sont conservées dans la pile.

3.3.5 Approche hiérarchique pour la traduction automatique

La traduction probabiliste à base de segment a pu résoudre certains problèmes de réordonnancement par rapport à une approche à base de mots notamment dans le traitement des expressions idiomatiques. Cependant, elle n'a toujours pas pu résoudre le problème lorsqu'il s'agit de réordonnancement très long qu'on peut rencontrer dans les paires de langues avec un ordre de mots assez différent (français-chinois par exemple).

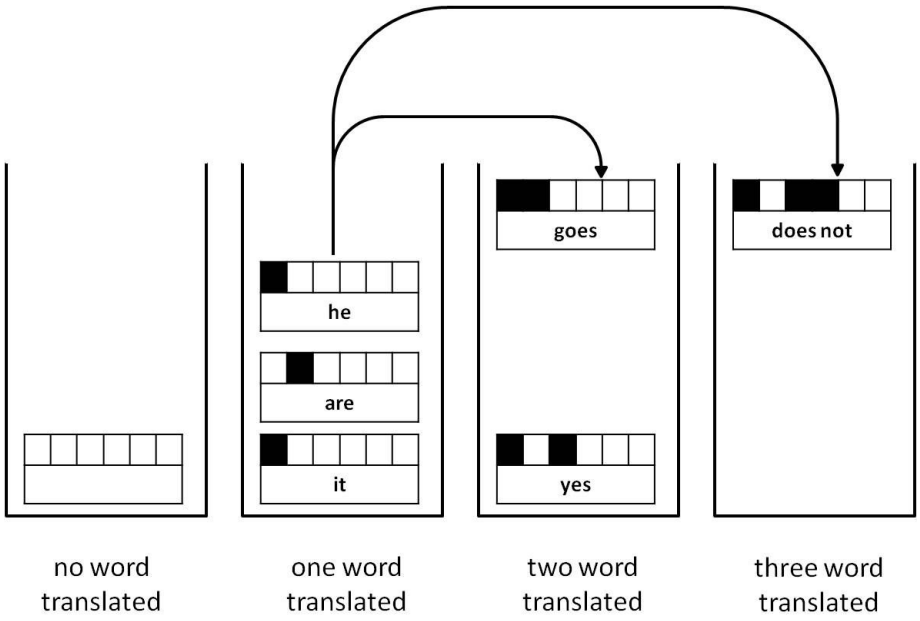

FIGURE 3.7 – *L'algorithme de décodage et les piles des hypothèses pour la recherche en faisceaux.*

Pour résoudre cette limitation (Chiang, 2007) a proposé l'utilisation de l'approche hiérarchique qui permet la résolution des problèmes difficiles du choix des ordres des séquences de mots traduits lors de la production de la traduction en langue cible.

L'idée derrière cette approche est d'utiliser des règles hiérarchiques (expressions) pour modéliser les réordonnancements sur la totalité de la phrase. Ces règles sont formalisées par des grammaires du type SCFG (synchronous context-free grammar) où chaque élément de cette grammaire est une règle de réordonnancement. Ces règles peuvent être apprises automatiquement à partir d'un corpus parallèle.

Une règle de réordonnancement peut être définie comme :

$$X \rightarrow < \gamma, \alpha, \sim >$$

où X est un symbole non-terminal, γ et α sont les séquences de mots qui contiennent des symboles non-terminaux et $\sim$ est l'alignement entre les symboles non terminaux dans γ et α.

Une règle de réordonnancement entre le français est l'anglais peut être par exemple :

$$\text{ne } X \text{ pas } \rightarrow \text{not } X$$

Ce modèle utilise les arbres comme structure de données. Donc la traduction consiste à produire un arbre dont les nœuds représentent la phrase cible à partir d'un arbre qui représente la phrase à traduire. Lorsqu'il n'est pas possible de traduire l'arbre en entier la traduction se fait sur des sous-arbres.

Le système de traduction hiérarchique peut être hybridé avec un autre système de traduction (à base de mots (Hayashi et al., 2010) ou à base de segment (Dymetman and Cancedda, 2010)). Pour l'instant, toutefois, ces modèles n'ont pas réussi à détrôner réellement les approches plus simples de type PB-SMT et font l'objet d'études assez nombreuses.

3.4 Outils pour la traduction automatique probabiliste

Aujourd'hui le développement d'un système de traduction automatique probabiliste est devenu relativement facile, grâce à la disponibilité de plusieurs corpus et outils libres pour apprendre les différents composants du système.

L'apprentissage du modèle de traduction nécessite un corpus parallèle bilingue aligné au niveau des phrases. Grâce aux sites ayant des versions multilingues, le web constitue une source importante pour collecter ce genre de corpus. Par ailleurs, plusieurs corpus sont disponibles grâce à la traduction de certains documents officiels (par exemple les actes du parlement européen (Koehn, 2005) ou ceux du parlement canadien (Simard, 1998)). Par ailleurs, la collecte des corpus pour apprendre un modèle de langage est rendue plus facile puisqu'il s'agit d'un corpus monolingue.

En 1990, l'outil "GIZA" a été développé pour apprendre un alignement en mots et implémenter des modèles IBM 1-3 (Al-onaizan et al., 1999). Une version plus avancée de cet outil, "GIZA++", à été développé par (Och, 2003a). Cette version implémente des modèles IBM-4 et HMM. Les modèles IBM-4 ont été aussi traités par l'outil "ISI ReWrite Decoder" développé par (Germann et al., 2001).

Le décodeur "Pharaoh" (Koehn, 2004) est le premier outil à implémenter des modèles de traduction à base de segments. Une autre boîte à outils, "MOSES", à été proposée plus récemment par (Koehn et al., 2007).

Cette boîte à outils regroupe un ensemble de scripts permettant de construire un système de traduction probabiliste à base de segments. MOSES contient le script GIZA++ pour aligner des mots, un script pour l'extraction de segments, un script pour apprendre et estimer les paramètres du système et également un décodeur.

MOSES implémente un modèle log-linéaire qui combine plusieurs modèles. Le décodeur est défini par défaut par 14 paramètres qui représentent les scores des différents modèles :

- 5 scores de modèle de traduction,
- 1 score de distorsion,
- 1 score de modèle de langage,
- 6 scores de modèle lexical,
- 1 score de pénalité de mots.

Les modèles de langage de type n-grammes peuvent être appris en utilisant la boîte à outils SRILM (Stolcke, 2002).

3.5 Evaluation des systèmes de traduction

En langage naturel, il n'y a pas une seule manière de traduire une phrase d'une langue vers un autre. L'évaluation d'une traduction automatique est une question non triviale du fait que plusieurs traductions peuvent être jugées correctes.

Distinguer une bonne traduction automatiquement n'est pas une question évidente. C'est pourquoi les tous premiers systèmes de traduction ont été évalués par des humains (évaluation subjective). Cette évaluation est coûteuse en temps et dépend énormément du jugement humain (et donc des humains qui les rendent) et peut ne pas être très homogène sur la totalité du corpus (surtout dans les cas où l'évaluation est faite par plusieurs personnes différentes).

Ces limites justifie le besoin d'adopter une évaluation automatique (objective) pour la traduction automatique. Cette évaluation nécessite un corpus bilingue dont la partie source servira en test et la partie cible servira en référence.

Le score TER

Le score TER (Translation Error Rate ou Translation Edit Rate) et un score d'évaluation de la traduction automatique basé sur le nombre minimum d'opérations à effectuer sur une hypothèse t_h de sortie de traduction pour la transformer en sa référence t_r. Cette métrique considère les opérations d'insertion, de suppression, de substitution et aussi de déplacement d'une séquence de mots (Snover et al., 2006). Un déplacement peut décaler toute une séquence de mots vers la droite ou la gauche et sera considéré comme une seule erreur. Le TER peut être donné comme suit :

$$TER(t_h = \frac{count(Ins) + count(Del) + count(Sub) + count(Dep)}{count(t_r)} * 100$$

Le score BLEU

Le score BLEU (BiLingual Evaluation Understudy) a été proposé par IBM (Papineni et al., 2002) pour l'évaluation objective de la traduction automatique. Le principe de ce score est la comparaison de la sortie du traducteur avec une/des traductions de référence. Les statistiques de co-occurrence et de n-grammes, basées sur les ensembles de n-grammes pour les segments de traduction et de référence, sont calculées pour chacun de ces segments et sommées sur tous les segments. Cette moyenne est multipliée par une pénalité de brièveté (pb), destinée à pénaliser les systèmes qui essaieraient d'augmenter artificiellement leurs scores en produisant des phrases délibérément courtes. Ce score peut être calculé comme suit :

$$BLEU = pb \,.\, \exp(\sum_{n=1}^{N} w_n . \log p_n)$$

sachant que N est la taille des n-grammes (souvent jusqu'à 4) et w_n sont des poids positifs tels que $\sum_{n=1}^{N} w_n = 1$. p_n qui représente le compte de précision de n-grammes sur la totalité du corpus

Le score BLEU est la métrique la plus souvent utilisée pour la traduction automatique. (Papineni et al., 2002; Doddington, 2002) ont montré que ce score a une bonne corrélation avec l'évaluation humaine. Malgré le fait que BLEU présente plusieurs limites (Zhou et al., 2006), il est considéré comme une méthode assez efficace pour des fins de comparaison de systèmes.

Le score NIST proposé dans (Doddington, 2002) part du même principe du score BLEU sauf que les n-grammes sont pondérés selon des fréquences d'apparition. Les n-grammes peu fréquents contribuent plus au score que les n-grammes fréquents.

Le score METEOR

Contrairement au score BLEU et NIST (qui sont des scores de précision), (Lavie and Denkowski, 2009) ont proposé le score METEOR (Metric for Evaluation of Translation with Explicit ORrdering) qui équilibre précision et rappel.

Le calcul de ce score est basé sur un alignement entre les uni-grammes de l'hypothèse et ceux de la référence. Un alignement peut être défini comme un ensemble de correspondances d'uni-grammes. Un uni-gramme d'une phrase est mis en correspondance avec zéro ou un seul uni-gramme d'une référence.

Les correspondances sont basées en premier lieu sur des formes orthographiques identiques, puis sur des mots de même racine (stemmisation) et enfin sur les synonymes. Cet alignement permet de mieux évaluer les hypothèses en tenant compte de plusieurs possibilités lexicales différentes de la référence. Le meilleur alignement est celui qui contient le plus grand nombre de correspondance d'uni-grammes avec le plus petit nombre de réordonnancements. Le score METEOR est déterminé à partir de ce meilleur alignement.

Le score METEOR (qui paraît assez efficace) dépend de la disponibilité des dictionnaires ce qui n'est pas évident dans toutes les langues.

3.6 Conclusion

Dans ce chapitre nous avons présenté un état de l'art des approches pour la traduction automatique. Les systèmes de traduction automatique peuvent être classés en plusieurs catégories selon leur architecture linguistique ou leur architecture computationnelle. Les systèmes de traduction automatique probabiliste ont connu une large évolution au fil des dernières années et ils ont l'avantage d'être assez performants et

faciles à implémenter grâce à la disponibilité de plusieurs outils libres pour les construire à partir d'un corpus bilingue parallèle.

Nous avons présenté aussi les différents composants d'un système de traduction probabiliste, notamment l'alignement des mots, le modèle de traduction et le modèle de langage. Enfin nous avons présenté des outils libres utilisés pour concevoir ces modèles notamment GIZA++, MOSES et SRILM et les métriques d'évaluation de la traduction automatique.

Deuxième partie

La portabilité multilingue d'un système de compréhension automatique de la parole

Chapitre 4

La portabilité d'un système de compréhension de la parole

Sommaire

4.1 Introduction 62
4.2 La portabilité des systèmes de dialogue 62
4.2.1 La portabilité des systèmes de reconnaissance automatique de la parole 63
4.2.2 La portabilité des systèmes de compréhension 64
4.3 Nos approches pour la portabilité multilingue d'un système de compréhension 66
4.3.1 La portabilité au niveau du décodage (TestOnSource) 67
4.3.2 La portabilité au niveau de l'apprentissage (TrainOnTarget) . . 68
4.4 La portabilité de l'annotation sémantique 69
4.4.1 Alignement direct (non-supervisé) 69
4.4.2 Alignement indirect (semi-supervisé) 70
4.4.3 Alignement obtenu pendant la traduction 71
4.5 Accroître la robustesse du système de compréhension aux erreurs de traduction 73
4.5.1 Apprentissage sur des données bruitées (SCTD) 74
4.5.2 Post-édition statistique (SPE) 75
4.6 Conclusion 76

4.1 Introduction

Aujourd'hui l'accessibilité aux services de technologies d'information à travers la parole reste d'une très grande importance pour une large catégorie de personnes pour laquelle c'est le seul moyen pratique d'accès à l'information. Donc pour un passage total vers la société numérique, le développement des services oraux ne doit pas être limité à certaines langues et à des domaines précis mais doit être généralisé pour couvrir un maximum de langues et d'applications possibles.

La généralisation des systèmes de dialogue oral accroît la nécessité du développement rapide de tels systèmes. Les systèmes de dialogue homme-machine peuvent être conçus pour différents domaines d'application et dans des langues différentes. La nécessité d'une production rapide de ces systèmes pour de nouvelles langues reste un problème ouvert et crucial auquel il est nécessaire d'apporter des solutions efficaces.

La portabilité des systèmes existant dans une langue donnée vers une nouvelle langue pour une tâche similaire est donc un enjeu important pour obtenir rapidement de nouveaux systèmes de dialogue.

Ce chapitre adresse la question du multilinguisme des systèmes de dialogue en présentant les travaux effectués dans ce domaine et nos propositions pour une portabilité multilingue du système de compréhension de la parole. Dans la section 4.2 nous introduisons d'une manière générale la portabilité multilingue de systèmes de dialogue, et nous présentons plus particulièrement dans la section 4.2.1 la portabilité du système de reconnaissance qui rejoint dans son esprit la portabilité du système de compréhension. Nous présentons les travaux dédiés à la portabilité de ce dernier (qui est le cœur de notre étude) dans la section 4.2.2. Les sections 4.3 et 4.4 présentent nos propositions pour la portabilité des systèmes de compréhension en utilisant des techniques de traduction automatique, alors que la section 4.5 présente nos propositions pour accroître la robustesse des approches proposées aux erreurs de traduction.

4.2 La portabilité des systèmes de dialogue

La construction d'un système de dialogue interactif soulève plusieurs problèmes liés au développement de ses différents composants (Glass, 1999). Cette procédure peut être coûteuse en temps et en expertise humaine. Le défi dans le développement d'un nouveau système de dialogue ne consiste pas uniquement dans la façon de construire les meilleurs modèles pour les différents composants du système, mais aussi à construire ces modèles tout en réduisant le coût et le temps du développement.

Dans les dix dernières années, de nombreux travaux ont traité la question du multilinguisme des systèmes de dialogue. Certains travaux ont proposé des outils et des plateformes pour faciliter la conception des composants principaux du système (Sutton et al., 1996), d'autres ont proposé de partir d'un système existant pour développer une nouvelle version dans une nouvelle langue (comme le cas du système de dialogue

Galaxy chinois (Seneff and Wang, 1997)) tout en bénéficiant des connaissances et des outils développés au cours de l'élaboration de la première version.

Certains composants du système de dialogue peuvent être considérés comme indépendants de la langue, par exemple, le gestionnaire de dialogue, et ne sont donc soumis à aucune question de portabilité à travers la langue (bien que cela puisse être remis en question par rapport à des habitudes culturelles qui peuvent influencer la façon dont les interactions sont traitées d'une langue à une autre (Burke et al., 2003)).

Cependant, d'autres modules, tels que la reconnaissance de la parole et le module de compréhension de la parole, diffèrent entre les langues et doivent être mis à jour pour porter le système de dialogue vers une nouvelle langue.

Etant donné que les modèles statistiques sont les plus utilisés récemment pour construire les composant majeurs des systèmes de dialogue et que la construction de ces modèles nécessite des données d'apprentissage, un nombre important des travaux qui ont traité le problème de la portabilité l'ont traité uniquement du point de vue de la minimisation de l'effort de collecte de nouvelles données dans une nouvelle langue (Glass et al., 1995; Gao et al., 2005; Fung and Schultz, 2008). Ces données sont ensuite utilisées pour apprendre les divers modèles (reconnaissance de la parole, compréhension, synthèse, ...).

Des travaux plus récents ont proposé des méthodes pour porter les différents modules du système de dialogue vers une nouvelle langue. Ces travaux sont résumés dans les sections suivantes. Il est important de noter que, malgré le fait que nous nous sommes intéressés plus particulièrement à la portabilité du système de compréhension de la parole dans le cadre de cette thèse, la portabilité des autres modules reste possible avec des approches similaires.

4.2.1 La portabilité des systèmes de reconnaissance automatique de la parole

De nombreuses études ont déjà été menées pour développer des méthodes rapides pour la portabilité d'un système de reconnaissance à travers les langues. Certains ont proposé de réduire l'effort manuel du développement du système de reconnaissance par l'utilisation d'un modèle générique (Lamel et al., 2001). D'autres ont suggéré de porter les composants du système de reconnaissance, aussi bien le modèle acoustique que le modèle de langage.

Certains travaux sont concentrés plus particulièrement sur la portabilité des modèles acoustiques. Ces travaux ont montré que le coût d'apprentissage d'un nouveau modèle acoustique peut être réduit en utilisant un modèle général et en proposant une adaptation de ces modèles en fonction de la langue cible (Schultz, 2004; Schultz and Black, 2006).

D'autres efforts ont été dédiés à la portabilité des modèles de langage (Kim and Khudanpur, 2003; Akbacak et al., 2005) notamment l'application de la traduction automatique et la recherche d'informations pour apprendre un modèle de langage dans

une nouvelle langue à partir d'une autre langue ou d'un autre domaine. Cependant, d'autres travaux ont été plus généraux en traitant la totalité du système de reconnaissance (modèle acoustique et modèle de langage) (Schultz, 2004; Lefèvre et al., 2005).

Il est important de mentionner que les approches élaborées pour la portabilité du module de synthèse sont comparables à celles du système de reconnaissance (voir par exemple le projet CMU SPICE (Schultz and Black, 2006)). Cependant, le générateur de langage naturel, étant la plupart du temps basé sur des règles, celui-ci est généralement porté vers une nouvelle langue grâce à une traduction manuelle de ses règles.

4.2.2 La portabilité des systèmes de compréhension

Le problème du multilinguisme des systèmes de compréhension est un sujet abordé de plus en plus ces dernières années. Plusieurs travaux ont traité le développement de systèmes de compréhension multilingues (par exemple (Minker, 1998; Siu and Meng, 1999; Komatani et al., 2001)). Ces travaux doivent être remis en cause pour deux raisons principales : d'une part les méthodes utilisées pour apprendre le module de compréhension ont récemment changé leur paradigme dominant (des méthodes à base de règles vers des approches statistiques fondées sur un étiqueteur de séquences) et d'autre part, le fait que les méthodes de traduction automatique sont désormais assez performantes et beaucoup plus faciles à développer et à adapter à un besoin précis. C'est le cas des approches telles que la traduction automatique statistique à base de segments (Phrase-Based Statistical Machine Translation, PB-SMT) (Koehn et al., 2003).

Malgré toutes les qualités des approches statistiques utilisées pour apprendre le système de compréhension, ces modèles nécessitent encore beaucoup de données pour bien fonctionner. La qualité de ces modèles dépend énormément de la quantité de données utilisées pour les apprendre.

L'influence de la taille de l'ensemble des données sur la performance du module de compréhension a déjà été étudiée (Bonneau-Maynard and Lefèvre, 2001). Dans la même ligne les auteurs de (Komatani et al., 2010) ont proposé un moyen de répartir les données de manière optimale pour apprendre plusieurs systèmes de compréhension en parallèle, selon des approches différentes. De ces études, nous pouvons conclure que, pour l'apprentissage d'un modèle statistique, un corpus annoté qui représente une couverture suffisante de la sémantique du domaine est nécessaire. Même si certains événements peuvent être sous-représentés, ils doivent apparaître au moins une fois pour les modèles afin de les prendre en compte (à son tour le modèle peut être amélioré par une étape suivante de re-apprentissage ou d'adaptation).

La portabilité de ces modèles vers une nouvelle langue consiste alors à porter la connaissance représentée par le corpus annoté, d'une langue à l'autre. Comme le montrent les auteurs de (Gao et al., 2005), la partie la plus coûteuse en temps pour construire un nouveau système de dialogue est de collecter, de transcrire et d'annoter des données pour le développement du module de compréhension. Par conséquent, pour réduire les coûts du développement d'un nouveau modèle, la portabilité doit permettre d'augmenter la productivité en terme de temps et d'effort humain.

Plusieurs suggestions peuvent être appliquées afin de minimiser le temps de la transcription et la collecte des données d'apprentissage. Alors que (Gao et al., 2005) et (Sarikaya, 2008) ont proposé de construire un mini corpus puis d'utiliser ce corpus pour construire un système pilote utilisé pour pré-annoter les collections de données successives, d'autres proposent un apprentissage actif pour réduire le temps requis pour l'annotation et la vérification de corpus (Tur et al., 2003, 2005).

Les recherches visant à construire un nouveau système de compréhension dans une nouvelle langue en utilisant la traduction automatique peuvent être divisées en deux grandes catégories. La première propose d'utiliser des techniques de traduction automatique pour transférer un système de compréhension existant de la langue source vers la langue cible. L'autre propose des méthodes pour aider les annotateurs humains afin d'accélérer la collecte de nouvelles données d'apprentissage. Ces approches peuvent être combinées : en premier lieu, un système de compréhension dans la langue cible est obtenue par la portabilité d'un système existant. Ce système est ensuite utilisé pour aider les annotateurs humains en leur fournissant des pré-annotations automatiques.

Cependant, cibler une catégorie ou une autre n'est pas strictement équivalent. Dans la première catégorie, le but est de pouvoir étiqueter des données dans une nouvelle langue, et donc ces approches sont intéressées particulièrement à dériver l'interprétation des entrées d'utilisateurs. Les approches de la deuxième catégorie cherchent à produire de nouvelles données d'apprentissage dans la langue cible et donc l'étiquetage de données (l'annotation) doit être plus exigeant (tel qu'inclure une annotation conceptuelle au niveau des mots et ne pas se contenter d'une annotation globale au niveau des phrases).

Le défi d'une portabilité de l'annotation est moins important lors de l'utilisation des méthodes de compréhension qui n'ont pas besoin d'annotation au niveau des mots. Dans (Lefèvre et al., 2010), l'utilisation des classifieurs sémantiques de tuple (Semantic Tuple Classifier, STC) est proposée. Ces modèles n'ont pas besoin d'informations d'alignement.

Récemment, plusieurs études de la première catégorie ont montré que l'utilisation de la traduction automatique à différents niveaux du processus de compréhension peut aider à porter le système de compréhension vers une nouvelle langue (Suendermann et al., 2009b; Servan et al., 2010; Lefèvre et al., 2010; Jabaian et al., 2010). Par exemple, dans (Suendermann et al., 2009b), les auteurs proposent de traduire automatiquement les données de la langue source vers la langue cible afin de réapprendre une grammaire stochastique pour effectuer la reconnaissance et l'interprétation dans la langue cible.

Une autre possibilité consiste à considérer la sémantique d'un domaine indépendante de la langue. Ensuite, une solution consiste à traduire le corpus d'apprentissage vers la langue cible et d'inférer les balises sémantiques dans le corpus traduit.

Comme décrit dans (Servan et al., 2010), les phrases du corpus d'apprentissage sont divisées en un ou plusieurs segments, chaque segment ayant une annotation sémantique. Traduire le corpus d'apprentissage en utilisant la segmentation associée permet

une mise en correspondance directe entre les segments traduits et des étiquettes sémantiques. (Servan et al., 2010) ont montré que la portabilité d'un système de compréhension utilisant cette approche est possible et donne une performance acceptable, que ce soit à partir de traductions manuelles ou automatiques.

4.3 Nos approches pour la portabilité multilingue d'un système de compréhension

Bien que nos propositions soient dans la lignée de ce qui vient d'être cité, l'originalité consiste à utiliser avantageusement certaines données existantes, notamment des donnée d'apprentissage traduites manuellement, sans ruiner nos efforts pour épargner la contribution humaine dans le processus de portabilité. Cela est possible à l'aide de la traduction automatique qui nous permettra d'apprendre des systèmes adaptés au domaine à un coût très faible.

Au cours de la dernière décennie, la qualité des systèmes de traduction automatique a été considérablement améliorée. Des approches statistiques sont maintenant très répandues et la disponibilité d'un certain nombre de boîtes à outils logiciels libres (MOSES, Josua, Jane...) permet de concevoir facilement des systèmes de traduction statistique. Une grande quantité de données peut être collectée sous forme de textes parallèles et ces données permettent d'apprendre des systèmes de traduction assez performants.

Cependant si des systèmes généraux ou des systèmes de large couverture sont disponibles pour la traduction automatique, leur performance ne reste guère satisfaisante pour les domaines spécialisés. Le développement récent de systèmes de traduction automatique statistique efficaces et faciles à utiliser est également d'une grande aide pour le développement rapide du système de traduction spécialisé dans le domaine, à condition que certaines données soient disponibles.

Contrairement aux corpus de traduction qui sont en général de l'ordre de centaines de milliers de phrases parallèles, il n'est guère possible de collecter plus de quelques milliers de dialogues, ce qui représente des dizaines de milliers d'énoncés (et d'ailleurs beaucoup d'entre eux sont très courts, comme des réponses oui-non).

Donc, en tenant compte de l'ordre de grandeur des corpus d'apprentissage disponibles en dialogue, nous ne sommes pas sûrs de la performance du système de traduction qui peut être obtenue. De plus les données sont monolingues pour commencer et donc une étape de traduction manuelle doit être effectuée avant que le processus automatique puisse avoir lieu. Il y a donc ici un problème du type "chicken and egg" : nous avons besoin de données du domaine (bitextes) pour apprendre un système de traduction spécialisé afin de générer de nouveaux bitextes à partir de données linguistiques source pour augmenter la taille des données d'apprentissage disponibles.

Dans la suite de ce document, la langue pour laquelle un système de compréhension existe déjà est appelée la langue source et la langue pour laquelle nous voulons

développer un nouveau système de compréhension est appelée la langue cible.

Comme déjà mentionné, dans la mesure où les approches statistiques sont ciblées, un corpus annoté sémantiquement est nécessaire pour former le module de compréhension. Par conséquent, afin de porter l'étiqueteur d'une langue vers l'autre, nous proposons plusieurs approches qui peuvent être divisées en deux grandes stratégies différentes selon le niveau où la portabilité est effectuée. L'application d'une stratégie ou d'une autre dépend de plusieurs facteurs notamment la tâche attribuée au système développé (insertion dans un système de dialogue ou utilisation pour faire de la pré-annotation), et la disponibilité des données qui ont servi à apprendre le modèle de base.

4.3.1 La portabilité au niveau du décodage (TestOnSource)

Dans de nombreux cas, le but de la portabilité d'un système de compréhension est de pouvoir étiqueter des entrées d'utilisateur dans une langue en utilisant un modèle existant pour une autre langue sans forcément vouloir obtenir un nouvel étiqueteur dans la langue cible.

Dans cette première approche, nous supposons qu'un système de compréhension est disponible en langue source, et nous utilisons un système de traduction automatique probabiliste pour traduire les entrées de l'utilisateur en langue cible vers la langue source. Ces traductions sont ensuite les entrées du système de compréhension original.

En d'autres termes, nous portons le système « au niveau du test » sans modifier le processus d'apprentissage de l'étiqueteur sémantique. Cette méthode sera nommée TestOnSource dans la suite de ce manuscrit.

Le système de traduction peut être obtenu en utilisant des méthodes standard (basées sur de grandes collections de textes parallèles), ou on peut vouloir développer un système de traduction adapté au domaine. Dans ce cas, nous proposons d'utiliser des données de dialogue traduites pour apprendre un système de traduction qui sera ensuite utilisé pour traduire l'ensemble de test. Un sous-ensemble du corpus d'apprentissage est traduit manuellement à partir de la langue source vers la langue cible, de manière à obtenir un corpus (petit mais très pertinent) parallèle.

Cette stratégie a l'avantage d'être très simple mais ses performances dépendront, bien évidemment des performances du système de traduction automatique utilisé pour revenir de la langue cible vers la langue source. La performance globale dépend aussi de la robustesse de l'étiqueteur sémantique aux erreurs de traduction (la robustesse du modèle de compréhension sera abordée plus tard dans cette thèse). Il est important de mentionner que — vu qu'une simple traduction des nouvelles entrées est suffisante pour la mise en place de cette méthode — cette méthode peut être appliquée parfaitement dans les cas où nous ne disposons pas des données qui ont servi pour apprendre l'étiqueteur sémantique (cas où le module de compréhension est une "boîte noire"). La méthode TestOnSource est résumée dans la FIGURE 4.1.

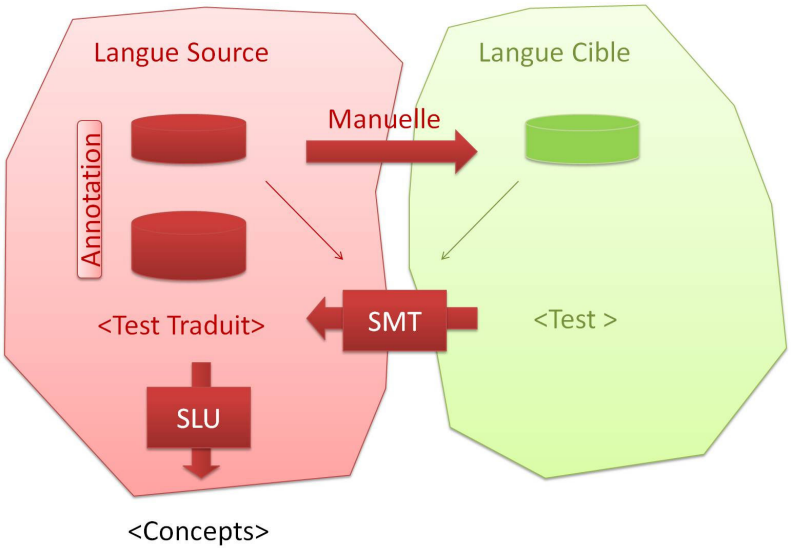

FIGURE 4.1 – *La méthode TestOnSource.*

4.3.2 La portabilité au niveau de l'apprentissage (TrainOnTarget)

Malgré le fait que la méthode TestOnSource (assez simple à appliquer) assure un étiquetage sémantique des entrées d'une nouvelle langue, le besoin d'un système de compréhension dans la langue cible reste indispensable pour certains cas notamment lorsqu'on cherche à obtenir un étiquetage au niveau des mots et non pas au niveau de la phrase. Cette seconde approche consiste donc à apprendre un nouveau modèle de compréhension dans la langue cible. L'idée générale derrière cette stratégie est de traduire le corpus d'apprentissage de la langue source vers la langue cible en premier lieu, puis d'inférer les annotations sémantiques correspondantes. En d'autres termes, cette stratégie consiste à porter le corpus d'apprentissage et son annotation pour apprendre un nouveau module de compréhension dans la langue cible. Cette méthode sera nommée TrainOnTarget dans la suite. TrainOnTarget est illustré dans la FIGURE 4.2.

Encore une fois, comme pour la méthode TestOnSource, nous proposons d'obtenir un corpus parallèle par la traduction manuelle d'un sous-ensemble des données d'apprentissage, en vue d'apprendre un système de traduction. Ce système pourra être utilisé ensuite pour traduire le reste du corpus d'apprentissage. Une fois celui-ci entièrement traduit, le défi est de porter l'annotation du corpus source vers le corpus cible. Ce transfert d'annotation est basé sur un alignement automatique entre les phrases sources et les phrases cibles. La portabilité de l'annotation d'un corpus existant vers sa traduction peut être faite en utilisant soit l'alignement direct soit l'indirect entre les deux corpus. Ceci peut également être fait pendant la traduction et non pas dans une étape séparée. La section suivante donne plus de détails sur ces différentes possibilités.

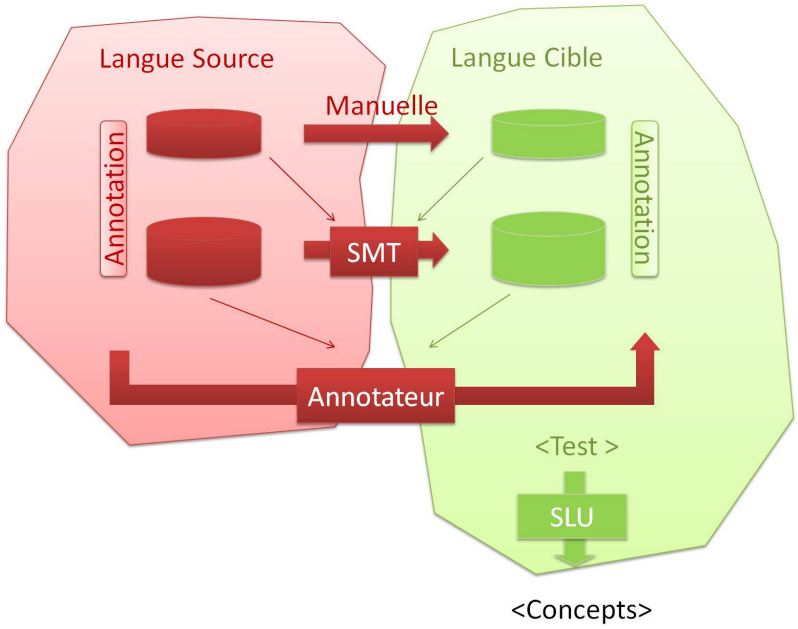

FIGURE 4.2 – *La méthode TrainOnTarget.*

4.4 La portabilité de l'annotation sémantique

Après avoir traduit le corpus d'apprentissage dans la langue cible, on a besoin de porter l'annotation du corpus source vers la version traduite. Une des raisons est que cette annotation est nécessaire au niveau des segments pour certaines approches statistiques de classification, comme les CRFs. Pour ce faire, plusieurs techniques de traduction automatique — telles que l'alignement au niveau mots entre les phrases source et les phrases cible — peuvent être appliquées. La projection de l'annotation vers le corpus de la langue cible peut alors être obtenue soit en mode direct soit en mode indirect, comme détaillé ci-dessous.

4.4.1 Alignement direct (non-supervisé)

L' alignement automatique de mots est une étape importante dans la traduction automatique statistique. Il consiste à générer des alignements entre les mots dans deux phrases ou séquences parallèles en se basant uniquement sur des informations statistiques extraites d'un corpus parallèle par un processus itératif de raffinement. Plusieurs outils sont disponibles pour l'alignement non supervisé de mots : GIZA++ (Och and Ney, 2000) qui utilise des modèles IBM et des modèles HMM, et aussi l'aligneur Berkeley (Liang et al., 2006) qui utilise une autre approche nommée "alignment by agreement". Ces modèles ont montré une bonne qualité d'alignement et sont utilisés largement pour construire des systèmes de traduction.

FIGURE 4.3 – *Exemple de projection d'étiquettes sémantiques en utilisant un alignement direct (non-supervisé).*

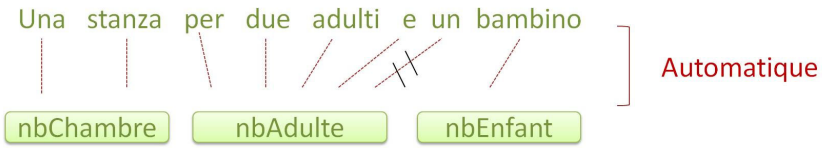

FIGURE 4.4 – *Exemple d'erreur en utilisant un alignement (non-supervisé) pour la projection de concepts.*

Afin de réduire le coût associé à la production d'un corpus annoté au niveau mots, (Huet and Lefèvre, 2011) ont proposé d'utiliser les informations de l'alignement direct entre une phrase et ses étiquettes sémantiques pour créer un corpus annoté au niveau segments. Un processus similaire peut être appliqué afin de porter l'annotation du corpus source à la traduction (cible). Pour cela nous proposons d'aligner automatiquement les phrases d'apprentissage traduites avec les séquences de concepts correspondant aux phrases source. Pour apprendre l'alignement, nous proposons d'utiliser un corpus parallèle composé de concepts provenant du corpus source d'un côté et les phrases obtenues par la traduction du corpus source de l'autre côté.

Cet alignement automatique nous permet de créer un corpus annoté, au niveau segmental, qui peut être utilisé plus tard pour apprendre un modèle de compréhension dans la langue cible. La FIGURE 4.3 donne un exemple d'un alignement automatique entre une phrase cible et les étiquettes sémantiques correspondantes. Bien que les alignements automatiques soient corrects dans la plupart des cas, les erreurs d'alignement ne sont pas négligeables pour certains autres cas, comme le montre la FIGURE 4.4. Pour améliorer cette situation, nous proposons de faire un meilleur usage de l'information disponible dans le corpus source dont nous disposons. Pour cela nous proposons aussi une méthode indirecte (semi-supervisée) pour l'alignement mots-concepts.

4.4.2 Alignement indirect (semi-supervisé)

Cette approche utilise l'alignement mot-à-mot entre les phrases source et les phrases cibles pour la projection des concepts sources vers les phrases cibles. Puisque le corpus d'apprentissage source est déjà annoté au niveau de segments, nous proposons d'utiliser l'information d'alignement mot-à-mot entre ce corpus et les corpus traduits pour aligner directement les concepts sémantiques avec les segments de la langue cible. Bien que pour de nombreux cas, cet alignement soit sans ambiguïté (voir FIGURE 4.5), cette

projection peut aussi conduire à des cas complexes, comme illustré dans la FIGURE 4.6.

Autrement dit, l'idée est d'utiliser les informations d'alignement mot-à-mot entre les phrases source et les phrases cible pour projeter l'alignement sémantique des segments de la langue cible aux segments de la langue source.

Le terme "semi-supervisé" pour cette méthode vient du fait qu'une intervention humaine a déjà eu lieu côté source pour annoter sémantiquement au niveau des mots le corpus d'apprentissage. Cette annotation est ensuite projetée automatiquement vers le nouveau corpus traduit.

Pour cela, un algorithme spécifique a été développé. Cet algorithme utilise les informations d'alignement et les frontières entre les segments dans le corpus source pour inférer des concepts dans le corpus cible. Pour résumer, pour chaque segment du corpus source, l'algorithme associe les mots correspondants du corpus cible en se basant sur les informations d'alignement automatique mot-à-mot entre les deux corpus. Ensuite il attribue les concepts associés à ces segments.

Dans la plupart des cas, le résultat d'une telle projection est correct, chaque mot des phrases de la langue cible est aligné avec un seul mot en langue source, ou avec plusieurs mots d'un seul segment. Dans certains autres cas, un mot de la langue cible peut être aligné avec deux mots qui appartiennent à deux segments différents ce qui peut créer une ambiguïté pour projeter les étiquettes. Afin de minimiser le bruit dans le corpus annoté créé, les mots de la langue cible alignés à différents mots du corpus source doivent être traités spécifiquement.

Chaque fois qu'un mot cible est aligné à plusieurs mots sources appartenant à des segments conceptuels différents (comme dans la FIGURE 4.6), l'algorithme doit décider à quel concept doit être associé le mot cible. Beaucoup de règles élaborées peuvent être introduites ici pour mettre en œuvre un processus de bonne décision, basé sur le contenu lexical et sémantique ou sur les informations de contexte. Cependant, pour mettre en place une méthode de base, notre proposition est la suivante : si un mot cible c_i est aligné à plusieurs mots source s_j, s_{j+1} appartenant à des segments conceptuels différents, nous ne prenons en compte que son alignement avec le mot qui apparaît en premier dans la phrase source s_j et donc le mot cible c_i appartiendra uniquement au segment conceptuel de s_j. Cette stratégie pourrait ne pas être la plus efficace ou ne donne pas la meilleure annotation dans tous les cas, mais elle a l'avantage d'être simple et cohérente sur l'ensemble des données d'apprentissage.

Cette stratégie permet d'annoter le corpus cible entièrement en utilisant l'information d'alignement en mots et l'étiquetage du corpus source.

4.4.3 Alignement obtenu pendant la traduction

Dans les propositions précédentes, la portabilité de l'annotation a été faite dans une étape distincte après la traduction du corpus d'apprentissage. Cette dernière proposition consiste à porter l'annotation en une seule étape qui combine la traduction du corpus source et le portage de son annotation.

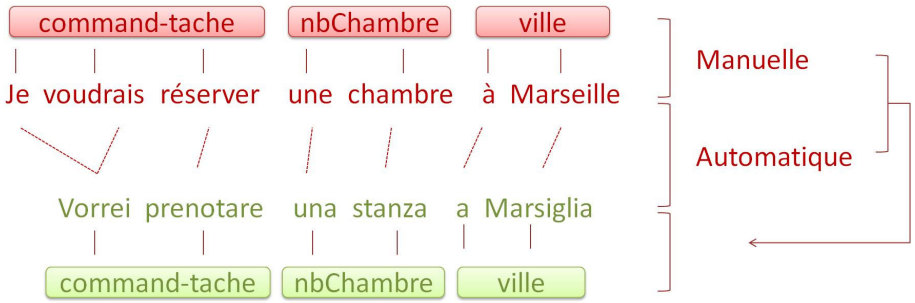

FIGURE 4.5 – *Exemple de projection de concepts sémantiques en utilisant un alignement indirect (semi-supervisé).*

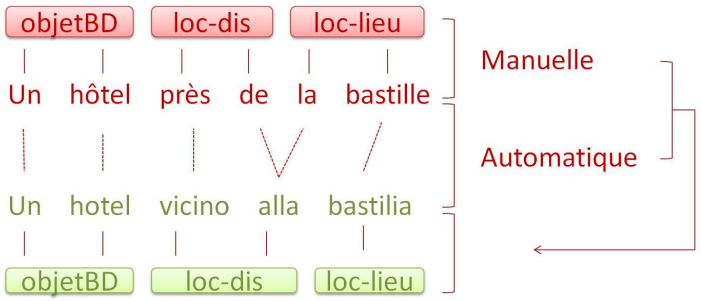

FIGURE 4.6 – *Exemple d'une situation ambiguë pour la projection de concepts.*

Dans cette approche, le corpus d'apprentissage est traduit en prenant en compte la segmentation en "segments sémantiques" (un "segment" est composé potentiellement de plusieurs mots mais correspond à une et une seule étiquette sémantique).

Pour dériver la représentation sémantique du corpus traduit, nous proposons de segmenter la phrase source avant de la traduire. En pratique cela consiste à utiliser une option du décodeur MOSES, décrit dans (Koehn et al., 2007), qui permet de forcer la segmentation conceptuelle d'une phrases avant la traduction. La segmentation des phrases est décrite par des balises XML dans le texte à traduire. Ces balises sont projetées dans le texte traduit, tout en empêchant toute tentative de modifier les frontières de segmentation de la langue source. En conséquence, nous obtenons le corpus d'apprentissage traduit en langue cible et annoté avec ses balises sémantiques correspondantes.

L'exemple donné précédemment peut alors être représenté sous la forme suivante :

```
<tag c=command_tache>
    Je voudrais réserver
</tag>

<tag c=nbChambre>
    une chambre
</tag>
<tag c=localisation_ville>
    à Marseille
</tag>
```

En utilisant l'option de MOSES qui prend en compte les tags XML comme information de segmentation, nous obtenons la sortie traduite suivante :

```
<tag c=command_tache>
    vorrei prenotare
</tag>
<tag c=nbChambre>
    una stanza
</tag>
<tag c=localisation_ville>
    a Marsiglia
</tag>
```

Tout le corpus d'apprentissage est traduit de cette façon avant un nouvel apprentissage du modèle de compréhension en langue cible.

4.5 Accroître la robustesse du système de compréhension aux erreurs de traduction

Nos expériences (qui seront présentées plus loin dans ce manuscrit), ainsi que d'autres travaux (Lefèvre et al., 2010; Jabaian et al., 2010), ont montré que la méthode la plus performante pour la portabilité d'un système de compréhension est aussi la plus simple, la méthode TestOnSource. Le défaut principal de cette méthode est que la qualité de l'étiquetage sémantique dépend principalement de la qualité de la traduction préalable. Ainsi, le système de compréhension doit prendre en compte des entrées bruitées par des erreurs de traduction.

Afin d'améliorer la robustesse de cette approche, nous proposons deux méthodes différentes qui peuvent être utilisées séparément ou mises en cascade. La première prend en compte le bruit venant de la traduction durant le processus d'apprentissage des modèles de compréhension ; la seconde corrige automatiquement la sortie du système de traduction avant de la transférer au système de compréhension. Il est important de noter que, bien que pas encore évaluées dans ce cadre, les deux méthodes sont aussi

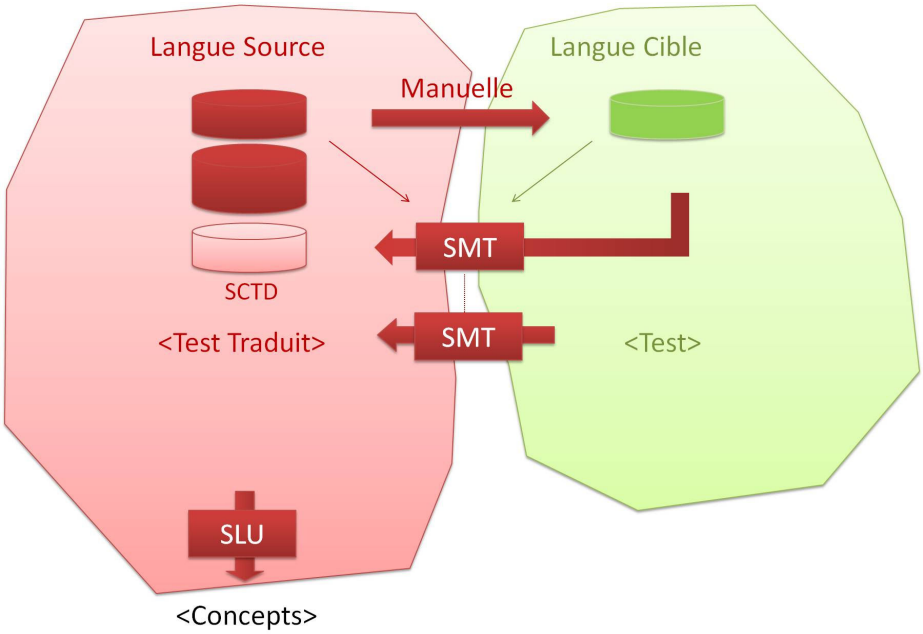

FIGURE 4.7 – *Accroître la robustesse de la méthode TestOnSource en utilisant des données d'apprentissage bruité.*

tout à fait adaptées pour traiter les erreurs dues à la reconnaissance automatique de la parole dans le cadre d'un système de dialogue réel.

4.5.1 Apprentissage sur des données bruitées (SCTD)

Vu que l'étiqueteur sémantique utilisé dans la méthode TestOnSource est appris sur des données propres (corpus source), sa capacité à étiqueter correctement des données bruitées (sortant d'un traducteur automatique) peut être limitée dans certains cas. Pour cela nous proposons une méthode d'apprentissage sur des données bruitées (Smeared Crosslingual Training Data, SCTD) dans laquelle des données similaires aux entrées automatiquement traduites sont prises en compte durant l'apprentissage du modèle.

Le principe de cette méthode est d'apprendre un modèle de compréhension (dans la langue source) avec des données additionnelles provenant de la sortie d'un système de traduction automatique.

En pratique, nous proposons de traduire les données d'apprentissage disponibles de la langue cible vers la langue source et ensuite d'inférer les concepts associés à ces données bruitées (en suivant la même méthode que TrainOnTarget). Puis nous ajoutons les données corrompues (maintenant annotées sémantiquement) aux données originales et l'ensemble est utilisé pour apprendre un nouveau modèle de compréhension (dans la langue source) qui alors intégrera le bruit présent dans les données traduites. Cette méthode est illustrée dans la FIGURE 4.7.

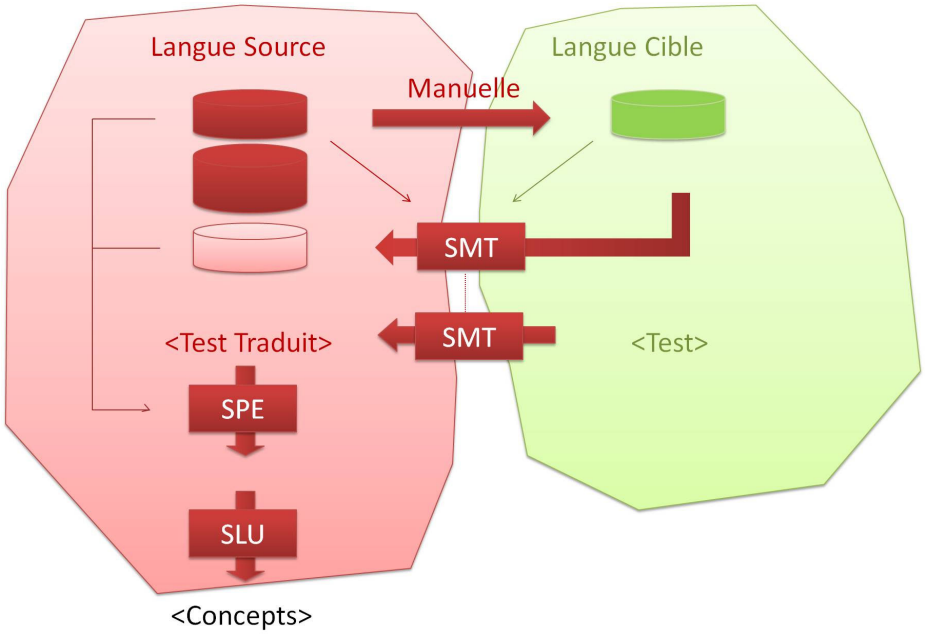

FIGURE 4.8 – *La mise en série des méthodes proposées pour accroître la robustesse de la méthode TestOnSource aux erreurs de traduction.*

4.5.2 Post-édition statistique (SPE)

Plusieurs travaux récents en traduction automatique comme (Simard et al., 2007; de Ilarraza et al., 2008; Béchara et al., 2011), ont proposé d'utiliser une approche basée sur un système de traduction pour post-éditer les sorties d'un autre système de traduction. Un tel système de post-édition statistique (Statistical Post Edition, SPE) a été proposé pour améliorer la qualité des données traduites avant leur envoi à des post-éditeurs humains. Pour entrainer un tel post-éditeur, (Simard et al., 2007; de Ilarraza et al., 2008) utilisent les sorties d'un système de traduction avec comme données parallèles leur post-édition manuelle.

Dans notre cas, dans la mesure où la sortie du système de traduction sera utilisée comme entrée du système de compréhension entrainé sur les données en langue source, nous proposons de post-éditer cette sortie afin de diminuer le bruit dû à la traduction.

Pour apprendre un post-éditeur statistique, notre choix a été de traduire automatiquement l'ensemble des données disponibles dans la langue cible, puis d'utiliser les sorties traduites avec leurs versions originales transcrites manuellement, comme corpus parallèle. Nous pensons que le module de post-édition permettra ainsi de réordonner quelques mots ou de retrouver des mots manquants dans un certain nombre de phrases. La FIGURE 4.8 illustre la mise en série des deux méthodes proposées pour la robustesse des systèmes de compréhension.

4.6 Conclusion

La portabilité d'un système de compréhension a pour but de minimiser le coût et l'effort humain lié à la création d'un nouveau système. Dans ce chapitre nous avons proposé plusieurs méthodes pour une portabilité rapide d'un système de compréhension de la parole vers une nouvelle langue. Ces méthodes, basées sur des techniques de traduction automatique, peuvent être classées dans deux catégories : la première propose de porter le système au niveau du test et donc traduire les entrées de l'utilisateur sans devoir apprendre un nouveau système ; la deuxième cherche à apprendre un nouveau système dans la nouvelle langue et donc porter les données d'apprentissage déjà disponibles vers la langue cible. Deux méthodes sont proposées pour accroître la robustesse du système de compréhension aux erreurs de traduction automatique. L'évaluation et la performance des méthodes proposées dans ce chapitre seront présentées dans le chapitre suivant.

Chapitre 5

Portabilité : expériences et résultats

Sommaire

5.1 Introduction 78
5.2 Matériau expérimental 78
5.2.1 Le corpus MEDIA 78
5.2.2 Les métriques d'évaluation 81
5.2.3 Les systèmes de traduction 81
5.3 Evaluation des approches proposées pour la portabilité 83
5.3.1 Les modèles de référence 83
5.3.2 Evaluation sur la totalité des données 84
5.3.3 Evaluation sur des données partielles 86
5.3.4 Evaluation des méthodes robustes aux erreurs de traduction . . 87
5.3.5 Combinaison 88
5.4 Validation des approches de portabilité proposées 89
5.4.1 Validation en utilisant des traductions en ligne 89
5.4.2 Validation sur une autre langue (arabe) 91
5.4.3 Pré-annotation automatique 93
5.5 Conclusion 96

5.1 Introduction

Le choix d'un couple de langues pour appliquer les méthodes proposées dans cette thèse dépend de considérations techniques et également des données disponibles. Disposer de données manuellement traduites ou annotées, disposer d'annotateurs ou d'outils spécifiques pour la langue cible, peut faire la différence quant au choix des langues. Dans cette thèse, nous avons proposé plusieurs approches pour la portabilité d'un système de compréhension automatique de la parole vers une nouvelle langue. Vu la disponibilité du corpus de dialogue MEDIA, la langue source est le français tandis que la langue cible considérée est l'italien puisque nous disposons au départ d'une partie du corpus MEDIA traduite manuellement en italien. La description des données disponibles ainsi que les outils et les métriques utilisés dans cette thèse se trouvent dans la section 5.2.

L'évaluation et la comparaison des approches proposées pour la portabilité ainsi que les approches proposées pour la robustesse des systèmes sont présentés dans la section 5.3. Nos premières évaluations des approches de portabilité du français vers l'italien supposent la disponibilité d'une partie des données françaises traduites manuellement en italien. Nous sommes conscients de la proximité des langues source et cibles dans cette étude et aussi de la difficulté à obtenir dans tous les cas des données cibles manuellement traduites, donc nous proposons une validation de ces approches une fois en utilisant des traductions en ligne et une autre fois en utilisant une langue cible différente de l'italien (l'arabe). Pour finir nous validons aussi nos propositions dans le cadre d'un scénario d'annotation semi-automatique réalisé dans le cadre du projet PORT-MEDIA 1.3. Les différentes validations sont présentées dans la section 5.4.

5.2 Matériau expérimental

Avant d'évaluer les approches proposées dans cette thèse, nous présentons le corpus MEDIA sur lequel nos expériences ont été réalisées, ainsi que les systèmes de traduction utilisés dans ces expériences et les différentes métriques d'évaluation utilisées.

5.2.1 Le corpus MEDIA

Toutes nos expériences sont fondées sur le corpus de dialogue français MEDIA. Le corpus MEDIA décrit dans (Bonneau-Maynard et al., 2005) couvre un domaine lié aux réservations d'hôtel et aux informations touristiques. Il s'agit d'une simulation d'un serveur téléphonique en français pour des réservations de chambre d'hôtel. Ce corpus est annoté par des étiquettes qui représentent la sémantique du domaine.

Le corpus est constitué de 1257 dialogues enregistrés par 250 locuteurs différents pour une durée totale de 70 heures d'enregistrement audio. Ces dialogues ont été collectés en utilisant un protocole de type "Magiciens d'Oz" (Wizard-of-Oz). Lors de l'enregistrement, un humain simule les réponses d'un serveur de dialogue, alors que l'util-

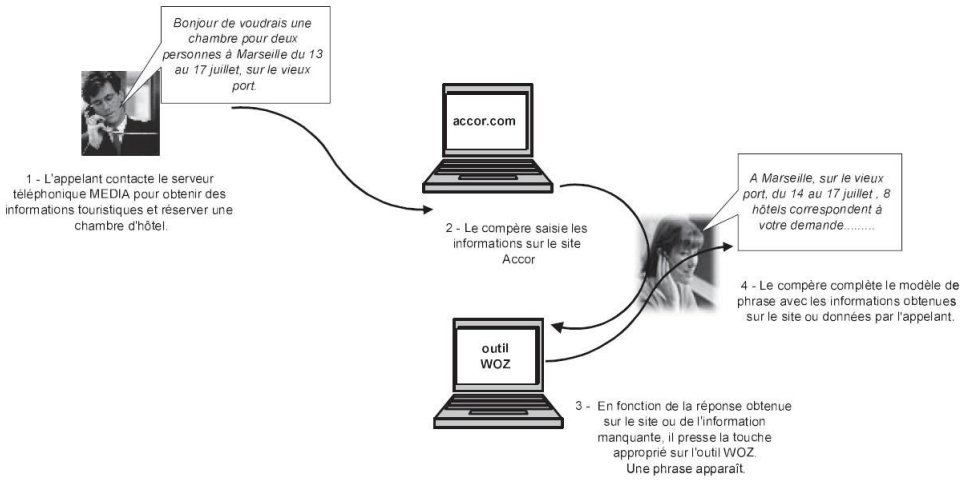

FIGURE 5.1 – *Protocole du Magicien d'Oz.*

isateur communique avec ce serveur. Le protocole de collecte des dialogues est illustré dans la FIGURE 5.1 (prise du "Manuel d'utilisation de l'outil WoZ. Projet MEDIA").

Les dialogues simulent des conversations téléphoniques pour des réservations touristiques. Les utilisateurs appellent le serveur pour réserver des chambres d'hôtel dans une ou plusieurs villes selon certains critères (distance du centre, période, équipement, prix, ...).

Les dialogues sont ensuite transcrits et annotés manuellement pour obtenir une représentation sémantique du domaine. La collecte et l'annotation du corpus ont été prises en charge par ELDA/ELRA [1]. L'outil semantizer (Bonneau-Maynard et al., 2005) est utilisé comme interface d'annotation du corpus MEDIA.

Les scénarios sont définis avec différents niveaux de complexité partant de réservations simples et requêtes bien précises jusqu'à des réservations multiples avec différents critères, des hésitations et des changements de plan au cours de la même discussion. Le tableau 5.1 présente un extrait de dialogue du corpus MEDIA.

Pour standardiser l'utilisation de ce corpus, les dialogues sont regroupés en 3 parties : un ensemble d'apprentissage (environ 13k phrases), un ensemble de développement (environ 1,3k phrases) et un ensemble d'évaluation (environ 3,5k phrases). Le tableau 5.2 donne une description détaillée des caractéristiques techniques du corpus MEDIA.

L'étiquetage sémantique du corpus est effectué par 99 concepts différents. Ces étiquettes peuvent être simples comme les dates ou les noms de villes ou peuvent être plus complexes comme les étiquettes linguistiques (celui-ci, le même, ...). Cette annotation sémantique découpe chaque phrase en plusieurs segments. Chaque segment cor-

1. http ://www.elda.org

Woz	Bienvenue sur le serveur MEDIA système d'information touristique et de réservations d'hôtels. Quelles informations désirez-vous ?
Locuteur	Je voudrais réserver un hôtel pour deux personnes à Paris
Woz	A Paris, quelle est votre date d'arrivée
Locuteur	Le huit juin
Woz	Pour combien de nuits souhaitez vous réserver ?
Locuteur	Deux nuits
Woz	Dans quel arrondissement souhaitez vous que votre hôtel se situe ?
Locuteur	Le sixième arrondissement
Woz	Trois hôtels correspondent à votre demande

TABLE 5.1 – *Un extrait de dialogue du corpus MEDIA.*

nombre de phrases utilisateur	18.801
nombre de phrases système	19.633
nombre de mots moyen par phrase utilisateur	8,3
nombre de mots moyen par phrase système	14,4
taille du vocabulaire utilisateur	2.715 mots
taille du vocabulaire système	1.932 mots
durée moyenne d'un dialogue	3,5 min

TABLE 5.2 – *Caractéristiques du corpus MEDIA*

respond à une unité sémantique. Ce segment est non seulement annoté par le concept lui-même mais aussi par une valeur (une normalisation du segment), une modalité (positive, négative, interrogative) et un specifieur qui représente la relation entre ce segment et les autres segments de la phrase. Un exemple d'annotation de MEDIA est donné dans le tableau 5.3.

Dans cet exemple le mode "+" signifie que la requête de l'utilisateur est affirmative. Cette précision permet de distinguer les modes des différents segments d'un énoncé et d'enlever l'ambiguïté dans des cas où ce n'est pas très clair, à partir d'un dialogue transcrit, si la phrase est affirmative ou interrogative.

Le spécifieur "réservation" permet d'obtenir une structure hiérarchique qui représente une réservation liée au concept "command-tache", alors que les valeurs représentent des valeurs normalisées correspondantes à la base de données. Il est important de mentionner que les expériences présentées dans cette thèse prennent en compte uniquement le concept et la modalité du segment.

Un sous-ensemble de données d'apprentissage (environ 5,6k phrases), de même que les ensembles de tests et de développement, ont été manuellement traduits en italien dans le contexte du projet européen LUNA (Servan et al., 2010).

	concept c	mode	spécifieur	valeur
Je voudrais réserver	commande-tache	+		réservation
un hôtel	objetBD	+	réservation	hotel
pour deux personnes	séjour-nbPersonne	+	réservation	2
à Paris	localisation-ville	+	hôtel	Paris

TABLE 5.3 – *Exemple d'annotation sémantique du corpus MEDIA.*

5.2.2 Les métriques d'évaluation

Dans nos expériences nous utilisons deux métriques d'évaluation différentes. La première métrique est utilisée pour évaluer la performance des moteurs de traduction utilisés pour la portabilité multilingue, alors que la seconde est appliquée pour évaluer les sorties de compréhension obtenues par les différentes approches.

Pour évaluer la traduction, nous proposons d'utiliser le score BLEU (voir section 3.5). Ce score est la métrique la plus souvent utilisée pour comparer les performances de différents systèmes de traduction. Par ailleurs nous proposons d'utiliser le CER (voir la section 2.5) pour évaluer les sorties des systèmes de compréhension.

Dans la mesure où seuls les concepts et les modalités sont pris en compte dans l'évaluation de l'étiquetage sémantique, la segmentation sémantique des phrases n'est pas prise en compte. Les valeurs normalisées sont généralement extraites par un système à base de règles. Notre étude étant dédiée à la portabilité des approches statistiques, ce système n'en fait pas partie et donc n'est pas considéré durant l'évaluation.

5.2.3 Les systèmes de traduction

Pour appliquer nos propositions de portabilité nous proposons d'utiliser des systèmes de traduction automatique spécifiques appris sur des données du domaine. Pour cela nous construisons deux systèmes de traduction pour obtenir des traductions du français vers l'italien et de l'italien vers le français. Pour réaliser ces systèmes, la boîte à outils MOSES (Koehn et al., 2007) est utilisée. MOSES implémente l'état de l'art des systèmes de traduction par segments utilisant des modèles log-linéaires.

La construction d'un système de traduction statistique nécessite un corpus parallèle pour apprendre le modèle de traduction et un corpus monolingue dans la langue cible pour apprendre le modèle de langage. Nous avons l'avantage d'avoir une partie de l'ensemble d'apprentissage du corpus MEDIA (5,6k phrases) traduite manuellement en italien et nous utilisons cette traduction manuelle comme corpus parallèle pour entrainer des modèles de traduction dans les deux directions. Chacune des parties séparément permet l'apprentissage d'un modèle de langage cible. L'ensemble de développement (1,3k phrases) avec sa traduction est aussi utilisé comme corpus parallèle pour ajuster les poids du modèle log-linéaire des systèmes appris (voir le tableau 5.5 pour un aperçu des ensembles de données traduites manuellement). Il est vraisemblable que

Système	FR –> IT	IT –> FR
MOSES	43,62%	47,18%
En ligne	39,75%	42,58%

TABLE 5.4 – *Evaluation des systèmes de traduction.*

Corpus MEDIA	Apprentissage	Développement	Test
MEDIA **français**	13K	1,3K	3,5K
Italien manuel	5,6K	1,3K	3,5K
Italien automatique	13K	-	-

TABLE 5.5 – *Aperçu du corpus* MEDIA *et de sa traduction vers l'italien (# phrases).*

la taille du corpus de développement est un peu petite pour MERT, mais nous avons souhaité garder la structure initiale du corpus.

Finalement, nous obtenons un système de traduction du français vers l'italien avec un score BLEU de 43,62% et un système de l'italien vers le français avec un score BLEU de 47,18%. Ces scores sont mesurés sur l'ensemble de test de MEDIA manuellement traduit (3,5k phrases). Dans la mesure où une seule référence par phrase est utilisée pour évaluer le score BLEU, de même que l'ensemble d'apprentissage est réduit, ces performances peuvent être considérées comme très acceptables comparées à d'autre tâches comme IWSLT (score autour de 30 % obtenue par des systèmes de traduction anglais-français appris sur un corpus de 30M phrases (Federico et al., 2012)). La performance de nos systèmes est comparée avec un traducteur automatique disponible en ligne (Google [2]) représentatif d'un système à large couverture. Les résultats de cette comparaison sont présentés dans le tableau 5.4. Confirmant notre hypothèse de départ, la comparaison montre que l'utilisation de données spécialisées pour apprendre des systèmes de traduction conduit à des systèmes plus performants que les systèmes génériques disponibles en ligne même si ce dernier fournit déjà des performances très correctes.

Le système de traduction français-italien est utilisé pour obtenir une traduction automatique de la partie restante (non traduite manuellement) du corpus d'apprentissage MEDIA. Ainsi une traduction intégrale (manuelle 5,6k + automatique 13K) est disponible. Le système italien-français est utilisé pour traduire le test italien en français, qui sera l'entrée du système de compréhension français comme proposé par l'approche TestOnSource. Le tableau 5.5 donne un aperçu des ensembles de données disponibles pour l'ensemble des expériences.

2. ces scores ont été obtenus en janvier 2011

Modèle	Sub	Del	Ins	CER
Français (13K)	3,1	8,1	1,8	12,9
Français (5,6K)	2,9	13,9	1,7	18,5

TABLE 5.6 – *Evaluation (CER %) du SLU français de référence.*

5.3 Evaluation des approches proposées pour la portabilité

Comme nous l'avons déjà mentionné nous avons choisi les CRFs comme méthode d'étiquetage sémantique pour nos expériences. Plusieurs outils, tels que CRF++ (Kudo, 2005) et Wapiti (Lavergne et al., 2010), permettent d'apprendre de tels modèles [3]. Dans un premier temps nous apprenons et évaluons notre modèle de référence (appris sur les données françaises de MEDIA), puis nous utilisons la traduction manuelle en italien des données de test afin d'évaluer les performances des approches. Premièrement nous évaluons et comparons les différentes approches proposées dans le chapitre précédent, nous évaluons ensuite nos propositions de robustesse aux erreurs de traduction et pour finir nous proposons de combiner les systèmes.

5.3.1 Les modèles de référence

La totalité de l'ensemble d'apprentissage de MEDIA est utilisée pour apprendre un étiqueteur français de base utilisant des fonctions unigrammes et bi-grammes. Cette base atteint de bonnes performances (12,9% CER) et peut être considérée comme une référence pour nos évaluations (performance comparable à des systèmes état de l'art conçus pour la même tâche (Hahn et al., 2010)).

Pour des comparaisons et des analyses que nous présenterons plus loin dans cette section, nous aurons besoin d'un modèle français appris sur une sous-partie des données disponibles. Un tel modèle a été appris sur le sous-ensemble de MEDIA pour lequel nous disposons d'une traduction manuelle (5,6K phrases). Ce modèle est bien sûr moins performant que celui appris sur la totalité des données (18,5% vs. 12,9%), mais sa performance reste acceptable par rapport à la taille de son corpus d'apprentissage. Les détails des erreurs de ces modèles sont présentés dans le tableau 5.6.

Une analyse des résultats montre que le pourcentage le plus important d'erreurs d'étiquetage vient des suppressions de concepts. Malgré une capacité à étiqueter correctement certains mots hors vocabulaire en se basant sur les relations du mot avec les concepts suivants et précédents, les CRFs restent sensibles aux mots hors vocabulaire et aux phrases mal formées provenant d'un dialogue téléphonique spontané. Les suppressions de lieux relatifs, noms des villes, des objets et noms d'hôtels sont fréquentes dans les hypothèses générées.

3. CRF++ est utilisé dans les expériences présentées dans la section 5.3 et Wapiti est utilisé dans les expériences présentées dans la section 5.4

FIGURE 5.2 – *L'hypothèse générée par le modèle CRFs pour la phrase "je voudrais réserver un hôtel à Nice" (le mot "Nice" est hors vocabulaire).*

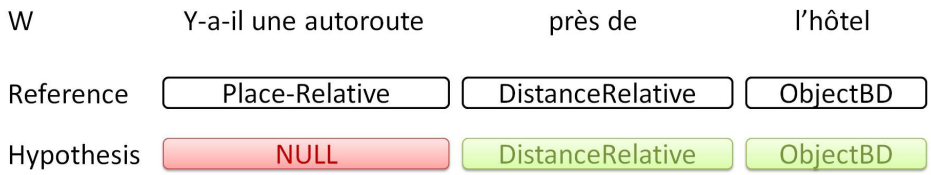

FIGURE 5.3 – *L'hypothèse générée par le modèle CRFs pour la phrase "Y-a-t il une autoroute près de l'hôtel" (le mot "autoroute" est hors vocabulaire).*

La FIGURE 5.2 montre un exemple d'une phrase étiquetée correctement par les CRFs malgré des mots hors vocabulaire (Nice est un nom hors vocabulaire dans cet exemple), alors que la FIGURE 5.3 montre un autre exemple où ce n'était pas le cas (autoroute est un nom hors vocabulaire dans cet exemple).

5.3.2 Evaluation sur la totalité des données

L'ensemble de test MEDIA traduit manuellement en italien est utilisé pour évaluer nos approches de portabilité. La traduction automatique de ce corpus (obtenu comme décrit dans la section 5.2.3) a été donnée comme entrée pour le système de compréhension de référence appris sur la totalité des données (décrit dans la section 5.3.1), ainsi nous obtenons les hypothèse générées par la méthode TestOnSource.

Nous avons également appliqué les trois méthodes TrainOnTarget telles que décrites dans la section 4.3.2.

Nous utilisons la traduction italienne de l'ensemble MEDIA (manuel + automatique) avec la version française comme corpus parallèle pour obtenir un alignement automatique mot à mots sur l'ensemble de l'apprentissage. Les informations d'alignements sont utilisées, comme décrit dans 4.4.1, et un modèle italien est appris sur les données obtenues pour évaluer la performance de la méthode d'alignement direct (nommée UnSupervisedAlignment par la suite).

Nous avons également aligné automatiquement chaque phrase italienne traduite directement avec la séquence de concepts correspondant à sa source française pour évaluer la méthode d'alignement indirect (décrite dans 4.4.2) (nommée Semi-SupervisedAlignment par la suite).

Modèle	Sub	Del	Ins	CER
SLU/CRF(TestOnSource)	5,2	12,1	2,6	19,9
SLU/CRF(TrainOnTarget)				
Un-SupervisedAlignment	3,6	15,3	2,5	21,4
Semi-SupervisedAlignment	3,1	15,0	2,3	20,5
TaggedTranslation	3,7	16,9	2,1	22,7

TABLE 5.7 – *Evaluation (CER %) des différentes stratégies de portabilité d'un système de compréhension de l'italien.*

Enfin, nous utilisons le système de traduction français-italien (décrit dans la section 5.2.3) pour traduire le corpus d'apprentissage segmenté par des balises XML correspondant à des segments sémantiques. Pour la suite, le modèle appris sur ces données sera nommée TaggedTranslation (décrit dans 4.4.3. Les performances obtenues par les différentes propositions sont comparées et présentées dans le tableau 5.7. Ces performances peuvent être comparées à la performance du modèle CRFs français appris sur la même quantité de données (19,9% CER).

Les résultats des trois approches montrent que la performance de la méthode de portabilité TestOnSource est légèrement meilleure que celle de la méthode TrainOnTarget, ce qui signifie que les CRFs sont plus robustes au bruit venant de la traduction automatique du test qu'au bruit dans les données d'apprentissage. Dans tous les cas, les résultats sont assez proches et peuvent être considérés comme bons et la dégradation de la performance en comparant à la référence française, est acceptable compte tenu de la réduction du temps nécessaire pour collecter de nouveau corpus pour l'apprentissage dans une nouvelle langue.

Il apparaît clairement à la lecture des résultats que la méthode Semi-SupervisedAlignment est la méthode la plus efficace parmi les méthodes TrainOnTarget. Ceci peut être expliqué par le fait que cette méthode est influencée uniquement par les erreurs d'alignement automatique, tandis que la méthode TaggedTranslation est influencée aussi par les erreurs de traduction automatique qui sont bien plus importantes. La semi supervision de cette méthode (par le corpus français) lui donne des avantages sur la méthode d'alignement direct.

Enfin, il est important de mentionner que pour les méthodes Un-SupervisedAlignment et Semi-SupervisedAlignment, une partie du corpus a été traduite manuellement tandis que l'autre partie a été obtenue automatiquement (par un système avec un score BLEU de 43,6%). D'autre part, la méthode TaggedTranslation utilise une traduction automatique de l'ensemble du corpus italien et le score BLEU du système de traduction italien-français biaisé par des segmentations XML est seulement 42,7%.

Les résultats montrent que (pour l'italien comme pour le français) le pourcentage le plus important d'erreurs d'étiquetage vient des concepts supprimés. Ces suppressions sont plus importantes pour l'italien que pour le français ce qui peut être expliqué par la traduction automatique incorrecte de certaines phrases que ce soit dans le test (TestOn-

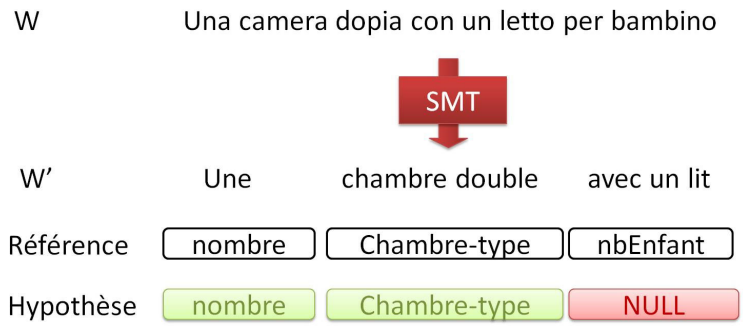

FIGURE 5.4 – *L'hypothèse générée par le modèle CRFs pour la phrase "una camera dopia con un letto per bambino" traduite par "une chambre double avec un lit"*

Modèle	Sub	Del	Ins	CER
SLU/CRF(TestOnSource)	5,1	15,2	2,8	23,1
SLU/CRF(TrainOnTarget)				
Un-SupervisedAlignment	4,6	17,2	1,7	23,4
Semi-SupervisedAlignment	4,7	16,7	1,5	22,8
TaggedTranslation	5,4	17,8	1,1	24,2

TABLE 5.8 – *Evaluation (CER %) des méthodes de portabilité appliquées sur 5.6k phrases traduites manuellement.*

Source) ou dans l'apprentissage (TrainOnTarget). De manière générale les concepts les plus concernés par la suppression sont les mêmes pour le français et pour l'italien (lieu relatif, noms des villes, objets, noms d'hôtels). La FIGURE 5.4 montre un exemple où la traduction imparfaite du test conduit à une suppression dans l'hypothèse produite par le modèle français ("un letto per bambino" a été traduite par "un lit" et donc l'information présente dans la traduction est incomplète).

5.3.3 Evaluation sur des données partielles

Afin d'évaluer l'influence de la taille et de la qualité des données d'apprentissage (traductions manuelles vs traductions automatiques) sur la performance du modèle de compréhension, nous proposons d'appliquer les méthodes de portabilité uniquement sur le sous-ensemble traduit manuellement (5.6K) des données d'apprentissage de MEDIA. Le modèle CRFs français appris sur le sous-ensemble français comparable (décrit dans 5.3.1) est utilisé comme référence de comparaison, ensuite les méthodes de portabilité sont appliquées. Les résultats sont rapportés dans le tableau 5.8. Ces performances peuvent être comparées à la performance du modèle CRFs français appris sur la même quantité de données (18,5(%) CER).

La dégradation de la performance du modèle français (12,9% vs 18,5%) a un effet direct sur la méthode TestOnSource puisque cette méthode implique directement ce

Modèle	Sub	Del	Ins	CER
TestOnSource	5,2	12,1	2,6	19,9
+SCTD	5,9	11,4	2,3	19,6
+SPE	6,5	10,6	2,5	19,7
+SCTD +SPE	6,4	9,9	2,9	19,3

TABLE 5.9 – *Evaluation (CER %) des approches proposées pour la robustesse des systèmes au bruit de traduction.*

modèle (3,2% CER absolu). L'impact des données supplémentaires bruitées est moins important pour les méthodes TrainOnTarget, mais toujours pas négligeable (environ 2% CER absolue).

Les données traduites automatiquement ont un double effet sur la méthode TrainOnTarget, d'une part elles augmentent la taille des données d'apprentissage et d'une autre part elles rajoutent du bruit à l'ensemble de données. Les résultats montrent que le gain de performance qu'on peut obtenir grâce aux données supplémentaires est plus important que la dégradation de performance due au bruit de traduction.

5.3.4 Evaluation des méthodes robustes aux erreurs de traduction

Nous avons tenté d'améliorer les performances de la méthode TestOnSource en renforçant sa robustesse aux erreurs de traduction. Pour cela nous évaluons les performances obtenues par nos deux suggestions présentées dans la section 4.5.

Pour cela nous avons traduit à nouveau automatiquement en français la partie italienne du corpus d'apprentissage (5,6K, voir tableau 5.5), de sorte à utiliser cette traduction en parallèle avec la partie originale (française également) pour entrainer le post-éditeur statistique (SPE) et fournir les données bruitées pour la méthode SCTD. Du point de vue de la performance de traduction exclusivement, l'utilisation de la post-édition automatique améliore le score BLEU du système de traduction de 47,18% à 49,25%.

Pour évaluer la méthode SCTD (décrite dans 4.5.1), nous apprenons un nouvel étiqueteur CRFs simultanément sur les données d'apprentissage françaises originales et traduites (+SCTD).

La méthode SPE (présentée dans la section 4.5.2), est aussi évaluée. Dans cette méthode le test traduit est post-édité et ensuite transmis au CRFs de français de référence (+SPE) ou CRFs appris par la méthode SCTD (+SCTD +SPE).

L'évaluation des performances de ces approches est donnée dans le tableau 5.9. Les deux méthodes, SCTD et SPE, améliorent les performances de l'étiqueteur sémantique. Leur mise en série donne les meilleures performances (19,3%) ce qui présente un gain de 0,3 absolu sur chaque méthode séparément.

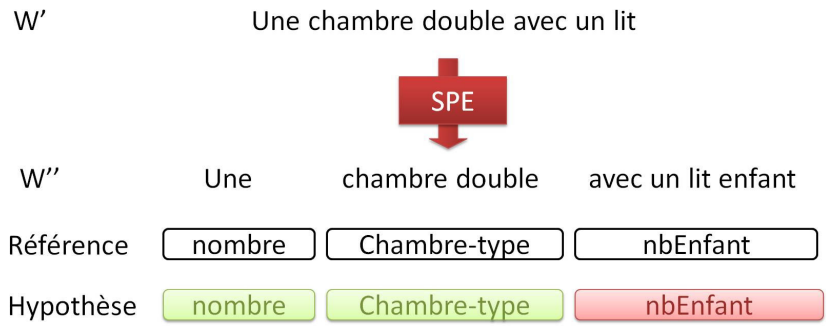

FIGURE 5.5 – *L'hypothèse générée par le modèle CRFs pour la phrase "una camera dopia con un letto per bambino" traduite par "une chambre double avec un lit" ensuite post-éditée par "une chambre double avec un lit enfant".*

La FIGURE 5.5 montre l'influence des approches de robustesse sur les hypothèses de sortie. La post-édition a permis de compléter la traduction automatique de la phrase et donc d'obtenir un étiquetage correct. Pour le même exemple présenté dans la FIGURE 5.4 où la séquence "un letto per bambino" a été traduite par "un lit" la post-édition permet de corriger cette traduction et de compléter l'information nécessaire pour un étiquetage correct de la phrase.

5.3.5 Combinaison

Après avoir évalué nos propositions de robustesse et vu que tous les résultats obtenus sont assez comparables, nous proposons de combiner les approches afin de bénéficier de leurs caractéristiques respectives pour améliorer la performance globale. La combinaison proposée est simple : un réseau de confusion est construit à partir des hypothèses des différents systèmes (Mangu et al., 2000) et la séquence de concept correspondant à la plus grande probabilité a posteriori est calculée. Pour cela nous utilisons l'outil "lattice-tool" de la boîte open-source SRILM (Stolcke, 2002).

Une première combinaison (appelée BASIQUE) consiste à combiner les méthodes de base (TestOnSource + les 3 approches TrainOnTarget). Dans un deuxième temps nous rajoutons les hypothèses produites par les approches de portabilité robuste au réseau de confusion. Cette combinaison est nommée TOUT par la suite.

Les résultats de ces combinaisons sont présentés dans le tableau 5.10. La légère amélioration de performance obtenue par la combinaison BASIQUE (-0,3% absolue) par rapport à la meilleure méthode reflète le fait que ces méthodes sont de nature assez similaire. Cette amélioration peut être plus importante lorsque qu'on intègre des mécanismes différents tels que le cas de la combinaison TOUT (-0.8% absolue) et cela encourage la combinaison avec des méthodes plus différentes (ce qui sera présenté dans la troisième partie de cette thèse).

Modèle	Sub	Del	Ins	CER
BASIQUE	6,5	10,6	2,5	19,6
TOUT	6,6	10,2	2,3	19,1

TABLE *5.10* – *combinaison de systèmes*

5.4 Validation des approches de portabilité proposées

Après avoir évalué et comparé les stratégies proposées pour la portabilité des systèmes de compréhension, nous cherchons à valider ces méthodologies. Une telle validation peut être effectuée à plusieurs niveaux. Tout d'abord, les techniques proposées dans ce travail font l'hypothèse qu'une partie des données, traduites manuellement de la langue source vers la langue cible, est disponible, permettant d'avoir un système de traduction spécialisé.

Maintenant dans la section 5.4.1, nous étudions l'utilisation directe d'un système de traduction généraliste " off-the-shelf" disponible gratuitement en ligne. L'utilisation d'un système de traduction disponible devrait nous permettre également de valider les approches proposées pour une autre langue avec une structure différente de l'italien, même si nous n'avons pas des données manuellement traduites dans cette langue. Le choix de la langue et les résultats expérimentaux sont décrits dans la section 5.4.2.

Dans le but d'évaluer l'utilisabilité de l'étiquetage sémantique obtenue par une approche de portabilité, la dernière validation consiste à utiliser les méthodes de portabilité afin d'aider les annotateurs humains en effectuant une pré-annotation automatique des données nouvellement collectées. Le processus d'annotation et les gains de productivité provenant de la pré-annotation sont détaillés dans la section 5.4.3.

5.4.1 Validation en utilisant des traductions en ligne

Les expériences présentées jusqu'à présent ne sont pas totalement non-supervisées. En effet, pour tous les cas nous avons utilisé des données traduites manuellement pour apprendre un système de traduction : soit pour traduire le test dans le cas d'application de la méthode TestOnSource soit pour compléter la traduction du corpus d'apprentissage et obtenir des informations d'alignement dans l'application des méthodes TrainOnTraget.

Le coût associé à cette traduction manuelle est relativement bas comparé à celui lié à la collecte et à l'annotation d'un nouveau corpus d'apprentissage, mais ce coût reste non négligeable. Dans cette expérience nous voulons vérifier qu'un tel coût est justifié par comparaison à une approche totalement non-supervisée et donc (potentiellement) « gratuite » [4].

4. bien sur de notre point de vue, car le développement d'un tel système a forcement requis des données d'apprentissage (sous une forme ou sous une autre)

	MOSES Translation				Online Translation			
Modèle	**Sub**	**Del**	**Ins**	**CER**	**Sub**	**Del**	**Ins**	**CER**
SLU/CRF(TestOnSource)	5,2	12,1	2,6	19,9	6,1	14,5	2,5	23,1
SLU/CRF(TrainOnTarget)								
Un-SupervisedAlignment	3,6	15,3	2,5	21,4	6,0	15,3	5,6	26,9
Semi-SupervisedAlignment	3,1	15,0	2,3	20,5	6,3	14,8	5,4	26,5
TaggedTranslation	3,7	16,9	2,1	22,7	5,5	15,4	5,7	26,6

TABLE *5.11* – *Evaluation (CER %) des différentes stratégies de portabilité en utilisant des traductions en ligne.*

Afin de répondre à cette interrogation nous proposons de reproduire les expériences en utilisant un système de traduction gratuit en ligne (généraliste) à la place de notre système de traduction appris pour la tâche spécifique.

Pour évaluer la méthode TestOnSource nous traduisons le test MEDIA en italien à l'aide d'une solution gratuite en ligne puis nous utilisons cette traduction comme entrée pour le modèle CRF de base. Pour la méthode TrainOnTarget deux approches ont été testées. Afin de permettre la comparaison avec la méthode TaggedTranslation, nous proposons de traduire les données d'apprentissage de MEDIA, segment par segment, en utilisant le traducteur en ligne, puis ces traductions sont associées aux étiquettes sémantiques initiales.

Dans un second temps, les données sont traduites intégralement en ligne et sont ensuite utilisées comme corpus parallèle pour l'approche d'alignement dans ses deux versions directes et indirectes.

Dans le cadre de nos expériences nous avons sélectionné Google comme traducteur en ligne. Le test MEDIA et sa traduction manuelle ont été utilisés comme couple test/référence dans chacune des directions de traduction. Pour l'italien vers le français ce traducteur donne un score BLEU 42,58% (à comparer à 47,18% obtenu par le système de traduction appris par nos soins sur les traductions manuelles), et pour le français vers l'italien de traducteur donne un score BLEU 39,75% (à comparer avec 43,62%). Les résultats de cette expérience sont reportés dans le tableau 5.11.

De manière attendue les performances des systèmes obtenus par cette approche non-supervisée sont inférieures à celles des systèmes utilisant une traduction manuelle d'une partie du corpus d'apprentissage. Toutefois malgré la dégradation du CER pour toutes les approches, son niveau absolu reste intéressant considérant les besoins de la tâche et la réduction substantielle du coût de développement.

La méthode Semi-SupervisedAlignment est la plus perturbée et devient presque équivalente à TaggedTranslation en version non-supervisée. Cela est dû au fait que dans cette version la méthode Semi-SupervisedAlignment ne prend plus l'avantage d'avoir une partie des données traduite manuellement et donc se met dans les mêmes conditions que TaggedTranslation. Le CER augmente de 22,7% à 26,6% (+3,9% absolu) pour TaggedTranslation et de 20,5% à 26,5% (+6%) pour l'alignement indirect.

TestOnSource perd 3,2% mais reste la plus performante tout en étant la plus facile et rapide à mettre en place. Ces résultats nous incitent à tester de nouvelles langues pour lesquelles nous ne disposons pas de traductions manuelles.

5.4.2 Validation sur une autre langue (arabe)

Nos expériences précédentes ont montré que la portabilité d'un système de compréhension du français vers l'italien peut être réalisée en utilisant des traductions en ligne avec une perte de 10,2% globale sur la performance du système (12,9 % pour le français à comparer avec 23,1 % pour l'italien, voir le tableau 5.12). Sachant que le français et l'italien sont des langues proches (avec un ancêtre commun, le latin), nous avons décidé de valider nos propositions de portabilité sur une autre langue avec une structure complètement différente. Pour cette tâche, l'arabe a été sélectionné. La disponibilité de certains volontaires parfaitement bilingues (arabe-français) nous a permis de traduire les énoncés du test MEDIA du français vers l'arabe.

La totalité du corpus de test de MEDIA a été traduite manuellement vers l'arabe par cinq volontaires vivant en France et ayant l'arabe comme langue maternelle. Nous avons utilisé l'ensemble du test arabe traduit automatiquement vers le français, afin d'évaluer la méthode TestOnSource. Cette méthode, qui donne la meilleure performance pour l'italien, peut être facilement appliquée à une nouvelle langue et sa performance dépend uniquement de la qualité de la traduction obtenue par le traducteur disponible.

En utilisant cette méthode pour la portabilité de l'arabe nous obtenons un système avec une performance (27,1 % CER) inférieure à celle obtenue pour l'italien (23,1 % CER). Cette différence de performance peut être expliquée par la différence de performance entre les traducteurs en ligne utilisés pour réaliser la traduction du test (score BLEU de 33,6% pour l'arabe à comparer à 42,6% pour l'italien [5]).

Contrairement à ce qu'on a pu obtenir pour la méthode TestOnSource, l'application directe des méthodes de TrainOnTarget n'est pas très efficace et rend la portabilité des systèmes en utilisant ces méthodes plus difficile pour l'arabe que pour l'italien.

Une des difficultés majeures rencontrée est la qualité de la traduction. Il est important de mentionner que le moteur de traduction utilisé pour cette expérience (Google) semble utiliser l'anglais comme langue pivot pour traduire du français vers l'arabe (qui n'est pas suffisant pour obtenir une qualité acceptable de traduction dans notre cas). Par exemple, le segment " Je voudrais réserver" sera traduit en arabe comme " Je voudrais un livre" ("réserver" traduit en anglais par "book" retraduit en arabe par livre). Ceci a une influence surtout lors de la traduction des phrases segment par segment pour la méthode TaggedTranslation. La qualité de la traduction obtenue est bien moins bonne que la traduction de la phrase complète. Ces segments traduits ne sont pas utilisables pour apprendre un nouveau modèle en arabe.

5. ces scores ont été obtenus en janvier 2011

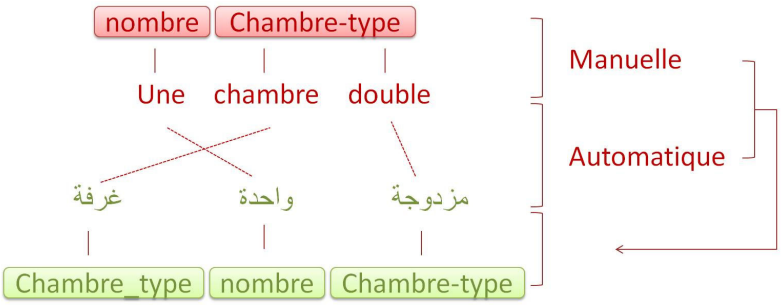

FIGURE 5.6 – *Exemple d'inadéquation en utilisant la méthode d'alignement.*

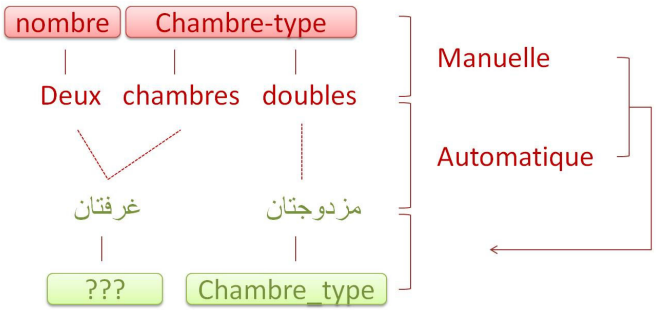

FIGURE 5.7 – *Exemple d'un alignement ambigu qui génère des erreurs sémantiques.*

Une autre difficulté importante pour l'application des autres méthodes du type TrainOnTarget est le fait que de nombreuses différences dans l'ordre des mots se produisent entre les phrases françaises et leurs traductions en arabe (contrairement à ce qu'on a pu observer pour l'italien) et donc les méthodes d'alignement ne sont pas utilisables sans un réordonancement des phrases arabes traduites. La FIGURE 5.6 représente un exemple d'un cas d'alignement qui conduit à une segmentation conceptuelle erronée. Un autre exemple (FIGURE 5.7) montre que dans certains cas, un mot composé (objet et le numéro en un seul mot) est aligné sur deux mots différents appartenant à deux concepts différents. L'utilisation de cet alignement sans pré-traitement conduira évidemment à des erreurs sémantiques.

Les exemples ci-dessus montrent le fait que l'obtention de corpus pour l'apprentissage d'un modèle de compréhension en arabe n'est pas vraiment possible en utilisant nos approches proposées sans pré-traitement des données traduites. Quoi qu'il en soit, la méthode TestOnSource est encore applicable pour la portabilité du système vers l'arabe malgré une perte de performance par rapport au système français de base. Le tableau 5.12 montre une comparaison entre les performances obtenues par la méthode TestOnSource pour les différentes langues et permet donc de juger de l'influence de la qualité de la traduction sur les performances des modules de compréhension.

Modèle	Sub	Del	Ins	CER
Français	3,1	8,1	1,8	12,9
IT-MOSES	5,2	12,1	2,6	19,9
IT-EnLigne	6,1	14,5	2,5	23,1
AR-EnLigne	7,3	16,9	2,9	27,1

TABLE 5.12 – *Evaluation (CER %) de la méthode TestOnSource pour l'italien et l'arabe.*

5.4.3 Pré-annotation automatique

Les expériences présentées précédemment permettent de comparer et d'évaluer la performance des stratégies proposées, mais que valent réellement les hypothèses sémantiques fournies par ces méthodes ?

Dans cette dernière expérience, nous proposons d'utiliser une méthode de portabilité pour effectuer une pré-annotation automatique d'un nouveau corpus. Cette pré-annotation sera ensuite corrigée manuellement par des annotateurs humains et les gains de productivité mesurés (un scénario d'annotation semi-supervisée). Ce gain de productivité pourra être considéré comme une mesure supplémentaire pour évaluer l'utilisabilité des approches de portabilité proposées dans cette thèse.

Pour effectuer la pré-annotation, nous devons premièrement choisir une méthode de portabilité à utiliser. Comme montré dans la section 5.3.2 les méthodes TestOnSource et TrainOnTarget sont assez efficaces, et leurs performances sont assez comparables, bien que la méthode TestOnSource donne des résultats un peu meilleurs que ceux obtenus par TrainOnTarget.

Toutefois, seule l'utilisation de la méthode TrainOnTarget permet de dériver directement une annotation au niveau des segments (basés sur les mots italiens). Comme la pré-annotation est concernée dans cette expérience et non pas seulement l'étiquetage, chaque séquence de mots du corpus doit être associée à une étiquette sémantique. La méthode TestOnSource annote la phrase mais pas les mots, ce qui donc rend la méthode TrainOnTarget plus utile pour ce processus d'annotation semi-supervisée.

De nouvelles données italiennes ont été collectées pour ce nouveau corpus. La collecte a été effectué exactement dans les mêmes conditions que la collecte du corpus MEDIA (tel que décrit dans la section 5.2.1), seule la langue des dialogues est différente. Des personnes natives italiennes ont été recrutées dans le cadre du projet Port-Media (Lefèvre et al., 2012). Un sous-ensemble des données transcrites a été automatiquement annoté par l'étiqueteur décrit ci-dessus. La pré-annotation du premier sous-ensemble de donnée a été corrigée manuellement par des annotateurs humains.

Les corrections ont été utilisées comme des données supplémentaires pour réapprendre l'étiqueteur sémantique et améliorer le système. Ensuite, un nouveau sous-ensemble de données a été automatiquement annoté avec le nouvel étiqueteur appris puis corrigé manuellement et renvoyé pour l'amélioration du système. Un ensemble de 604 dialogues a été annoté ainsi en trois itérations selon le processus décrit ci-dessus.

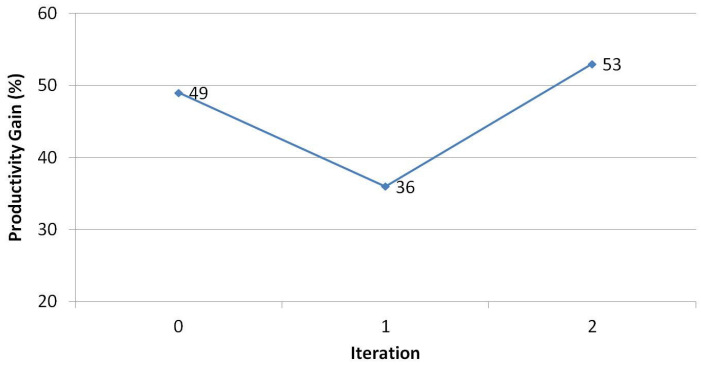

FIGURE 5.8 – *Pourcentage de gain de productivité de l'annotation sémantique.*

Iter.	Prod. Gain	Modèle	CER
0	49	MEDIA	20,8
1	36	+BLOC00	18,7
2	53	+BLOC01	16,4

TABLE 5.13 – *Le gain de productivité% et le CER% pour chaque itération du processus d'annotation sémantique. L'ensemble de test* PMEDIA *utilisé pour mesurer les CER a été compilé après le processus d'annotation avec des données de chaque bloc.*

Durant les itérations de l'annotation sémantique, nous avons constamment calculé les gains de productivité obtenus avec la pré-annotation. Pour cela, à chaque itération, un sous-ensemble de 10 dialogues a été annoté par deux annotateurs différents : l'un corrigeait les pré-annotations, tandis que l'autre annotait en partant de la phrase en italien seule.

Un gain de productivité peut être estimé par la différence entre le temps requis pour l'annotation et celui requis pour la correction du même sous-ensemble. La procédure de pré-annotation conduit à des gains de productivité de plus de 50% pour la troisième itération du processus d'annotation. La FIGURE 5.8 donne les gains de productivité pour cette procédure d'annotation semi-supervisée.

Nous avons également évalué le modèle utilisé pour la pré-annotation à chaque itération sur le nouvel ensemble de test PMEDIA. Les résultats de cette évaluation sont présentés dans le tableau 5.13. L'influence des données supplémentaires corrigées peut être vue par la diminution du CER d'une itération à une autre. Il est important de mentionner que l'ensemble de test PMEDIA a été extrait ultérieurement à partir des différent blocs.

Nous avons noté que le gain de productivité pour l'itération 1 est moins important que celui de l'itération 0 malgré l'amélioration du modèle utilisé pour la pré-annotation (passant de 20,8% à 18,7%) mesurée sur l'ensemble de test final. Pour clarifier cela, nous avons comparé la complexité des différents blocs en les étiquetant par le même modèle.

Iter.	Modèle	Test	CER
0	MEDIA	BLOC00	20,9
		BLOC01	21,9
		BLOC02	21,1

TABLE 5.14 – *Les CER(%) obtenues par le modèle* MEDIA *sur les différents blocs d'annotation du corpus* PMEDIA.

Le CER obtenu par le modèle de base utilisé pour la pré-annotation (nommée modèle MEDIA par la suite) sur les différents blocs (représenté dans le tableau 5.14), montre que le bloc01 est un peu plus complexe que les autres (21,9 % vs 20,9% pour bloc01 et 21,1 % pour bloc02) ce qui peut expliquer que le gain de productivité diminue dans la seconde itération (de 49% à 36%).

Le nouvel ensemble de données, dénommé PMEDIA, contient 604 dialogues. Le corpus est divisé en trois sous-ensembles (apprentissage, développement et test). 200 dialogues ont été sélectionnés au hasard pour former l'ensemble de test. Cette quantité (équivalente au test MEDIA) est suffisante pour une évaluation significative du modèle.

Les données d'apprentissage ont été utilisées pour apprendre un nouvel étiqueteur sémantique (PMEDIA), qui est évalué sur deux ensembles de test : le premier est le corpus de test traduit manuellement de MEDIA (corpus utilisé précédemment pour les expériences rapportées dans cette thèse), et l'autre est le nouvel ensemble de test collecté (PMEDIA). Nous évaluons également l'étiqueteur sémantique utilisé pour la pré-annotation (nommée modèle MEDIA DANS LA SUITE) sur les deux ensembles de tests.

Pour finir, les deux corpus MEDIA et PMEDIA ont été fusionnés pour apprendre un nouveau modèle combiné sur l'ensemble de données disponibles. Cette fusion permet d'augmenter la taille du vocabulaire et des données d'apprentissage. Ce modèle a été également évalué sur les deux corpus de tests.

Les résultats de ces évaluations sont présentés dans le tableau 5.15. Les résultats de l'évaluation sur le modèle MEDIA montrent la robustesse de ce modèle. Les performances de ce modèle sur les deux ensembles de test sont très proches (20,5% vs 20,8%). Ce modèle est autant performant sur un nouveau corpus de test que sur un corpus de test traduit.

Le modèle PMEDIA donne un CER de 18,9% sur le corpus de test qui lui est associé, tandis que sa performance est plus faible sur l'ensemble de test MEDIA. Cette différence de performance peut être expliquée par une différence dans la couverture conceptuelle entre le corpus MEDIA et le corpus PMEDIA.

La combinaison des deux corpus donne un modèle de meilleure qualité que chacun d'eux séparément, aussi bien sur le corpus MEDIA (19,5%) que sur le corpus PMEDIA (17,6%). On peut cependant noter que l'amélioration des performances reste limitée et cela peut être dû au fait que les deux corpus sont assez similaires.

Modèle	Test	Sub	Del	Ins	CER
MEDIA	MEDIA	3,1	15,0	2,3	20,5
	PMEDIA	3,8	13,9	3,1	20,8
PMEDIA	MEDIA	4,7	17,4	3,2	25,3
	PMEDIA	3,6	12,1	3,3	18,9
combinaison	MEDIA	2,8	14,6	2,1	19,5
	PMEDIA	3,9	9,0	4,6	17,6

TABLE *5.15* – *Evaluation (CER %) des modèles italiens sur deux ensembles de test différents.*

5.5 Conclusion

Dans ce chapitre nous avons évalué nos méthodes pour la portabilité d'un système de compréhension de la parole d'une langue vers une autre. Nos expériences reposent sur le corpus français MEDIA et nos propositions de portabilité ont été évaluées sur la portabilité de la compréhension du vers l'italien. Nous avons comparé deux méthodes principales TestOnSource et TrainOnTarget.

Nos propositions ont été évaluées en deux modes : le premier semi-supervisé dans lequel nous avons l'avantage de disposer de données traduites manuellement pour apprendre un traducteur automatique qui sera ensuite utilisé pour utiliser pour réaliser nos propositions de portabilité, tandis que le deuxième est totalement non-supervisé et utilise des traducteurs en ligne pour réaliser la traduction automatique.

Nos expériences ont montré que la méthode TestOnSource est plus performante que les méthodes TrainOnTarget que ce soit en utilisant un traducteur automatique spécialisé ou un traducteur générique en ligne.

Cette méthode a l'avantage de pouvoir être facilement et directement applicable sur toutes paires de langues à partir du moment où nous disposons d'un système de traduction entre ces deux langues.

Nous avons montré que les performances de cette méthode dépendent de la qualité du système de traduction automatique utilisé. Les performances de cette approche peuvent être améliorées en utilisant nos propositions de robustesse aux erreurs de traduction.

Nous avons montré aussi que la méthode TestOnSource a des performances acceptables même pour la portabilité vers une langue plus éloignée comme l'arabe.

Enfin nous avons proposé d'utiliser les méthodes de portabilité pour pré-annoter un nouveau corpus avant de le faire corriger par des annotateurs humains et nous avons montré que cette pré-annotation permet d'obtenir des gains importants de productivité et de diminuer considérablement le temps d'annotation de corpus.

Dans une expérience comparable nous avons montré que la notion de portabilité ne s'applique pas uniquement sur la portabilité multilingue mais aussi sur la portabilité

vers des nouveaux domaines. Le corpus MEDIA a servi pour apprendre un système de compréhension qui a été utilisé pour la pré-annotation d'un corpus pour un nouveau domaine. Plus de détails sur cette expérience sont présentés dans l'annexe 1.

Troisième partie

Approches conjointes pour la traduction et la compréhension

Chapitre 6

Approche générative vs. approche discriminante pour la traduction et la compréhension de la parole

Sommaire

6.1 Introduction 102
6.2 Méthode de traduction pour la compréhension 103
6.2.1 Adaptation des méthodes 104
6.2.2 Application à la portabilité multilingue 106
6.3 Méthode de compréhension pour la traduction 106
6.3.1 Modèle du LIMSI (FST/CRF) 107
6.4 Décodage conjoint pour la traduction et la compréhension, cas de la méthode de portabilité TestOnSource 112
6.5 Conclusion 114

6.1 Introduction

Aujourd'hui, les approches probabilistes sont les plus utilisées pour toutes les applications du traitement automatique de la langue (traduction automatique, étiquetage sémantique, reconnaissance de la parole...). La performance d'une approche dépend énormément de la tâche pour laquelle elle est utilisée et, selon les tâches, les approches permettant les meilleures performances ne sont pas toujours les mêmes.

Par exemple, pour une tâche de compréhension de la parole, les modèles discriminants, notamment les CRFs, sont les plus performants (Hahn et al., 2010), alors que pour la traduction automatique, ce sont les modèles génératifs, notamment les PB-SMT, qui sont le plus souvent utilisés.

Cependant, malgré les différences entre les approches probabilistes, celles-ci présentent des points communs et les frontières entre les unes et les autres ont tendance à s'estomper. On voit, par exemple, des travaux autour de l'utilisation d'approches discriminantes pour la traduction automatique (Och and Ney, 2002; Liang et al., 2006; Lavergne et al., 2011), tandis que les approches à bases de segments, rendues populaires par l'outil MOSES, ont tendance à être utilisées dans d'autres tâches du traitement automatique de la langue : conversion graphème-phonèmes (Rama et al., 2009), Part-Of-Specch (Gascó i Mora and Sánchez Peiró, 2007)...

Dans cette partie de la thèse nous cherchons à évaluer et comparer une approche discriminante (CRFs) et une approche générative (PB-SMT), aussi bien pour une tâche de compréhension que pour une tâche de traduction. Pour cela nous proposons d'utiliser et d'optimiser une approche PB-SMT pour la compréhension de la parole, et aussi de construire des modèles à base de CRFs pour la traduction automatique.

Cette étude à un double objectif :

- d'une part elle nous permettra de démontrer les **spécificités de chaque tâche** et d'évaluer les performances des approches respectives sur ces tâches, et
- d'autre part nous postulons que les approches sont assez complémentaires et donc **la combinaison de leurs sorties** devrait permettre d'augmenter la performance globale obtenue.

Enfin, vu que la portabilité multilingue est l'objet principal de cette thèse, et que (comme on l'a montré dans la deuxième partie de ce manuscrit) cette portabilité est obtenue en mettant un module de traduction en cascade avec un module de compréhension, nous proposons de réaliser un décodage conjoint entre les modules.

Dans certains cas, la meilleure hypothèse de traduction n'est pas l'hypothèse pour laquelle le système de compréhension génère la meilleure hypothèse (souvent pour des raisons d'ordre de mots) et donc la sélection préalable de la meilleure traduction n'optimise pas forcément le système lorsqu'on se place selon un scénario de portabilité de système.

Un décodage conjoint permettra de sélectionner des traductions en tenant compte de l'étiquetage qu'on pourra obtenir sur ces traductions. Dans cet esprit, nous ne cher-

chons plus la meilleure traduction possible mais la traduction qui sera étiquetée sémantiquement de la meilleure manière possible.

Ce chapitre est organisé de la manière suivante : la section 6.2 présente l'utilisation des techniques de traduction automatique pour la compréhension de la parole et propose également des adaptations itératives qui peuvent être appliquées pour prendre en compte les caractéristiques de la tâche. La section 6.3 décrit l'utilisation des CRFs pour la traduction automatique, ainsi que le modèle de traduction combinant FST et CRFs proposé par le LIMSI. Enfin, notre proposition pour un décodage conjoint entre la compréhension et la traduction est présentée dans la section 6.4.

6.2 Méthode de traduction pour la compréhension

Le problème de la compréhension d'un énoncé utilisateur peut être vu comme un problème de traduction de la séquence de mots qui forme cet énoncé (langue source) vers une séquence de concepts (langue cible). (Macherey et al., 2001, 2009) ont montré que les approches de la traduction automatique statistique peuvent être utilisées avec succès pour une tâche de compréhension de la parole.

Cette approche part du principe que les séquences de concepts sont les traductions des séquences de mots initiaux, ainsi la meilleure séquence de concepts c à partir d'une séquence de mots s est définie par :

$$\hat{c} = \text{argmax}_c P(s \mid c).P(c)$$

Pour résoudre cette équation sont requis :
- un modèle de langage de concepts $P(c)$ (qui représente la probabilité d'apparition d'un concept ou d'une suite de concepts sémantiques), et
- un modèle de traduction $P(s \mid c)$.

L'utilisation d'une approche de compréhension telle que les CRFs, nécessite un corpus d'apprentissage annoté au niveau des mots. Afin de minimiser le coût lié à cette annotation, des travaux ont proposé d'utiliser des modèles IBM pour obtenir automatiquement cette annotation à partir d'un corpus parallèle ("mots | | concepts") (Huet and Lefèvre, 2011).

L'alignement automatique est une étape indispensable pour apprendre des modèles de traduction à partir de corpus parallèles. Donc l'utilisation d'une approche de traduction pour la compréhension permet d'éviter de recourir à un corpus d'apprentissage annoté au niveau mot, car elle nécessite alors uniquement un alignement au niveau des phrases.

L'outil SRILM (Stolcke, 2002), peut être utilisé pour apprendre un modèle de langage n-grammes à partir de corpus de séquence de concepts, et la boîte à outils MOSES (Koehn et al., 2007) peut être utilisée pour apprendre un modèle de traduction PB-SMT à partir d'un corpus parallèle "phrases | | concepts". Le modèle log-linéaire implémenté par défaut dans MOSES est défini par 14 paramètres (voir section 3.4). Ces poids

peuvent être optimisés par un apprentissage à taux d'erreur minimum (Minimum Error Rate Training, MERT) qui est traditionnellement utilisé pour optimiser le score BLEU.

6.2.1 Adaptation des méthodes

Malgré le fait que cette approche suppose que la tâche de compréhension d'une phrase est une tâche de traduction de cette phrase vers des concepts, la compréhension a ses spécificités qui doivent être prises en considération afin de pouvoir améliorer les performances obtenues par cette approche.

Les différences entre une tâche de traduction classique (d'une langue vers une autre) et l'utilisation de la traduction pour la compréhension (traduction d'une langue vers des étiquettes sémantiques) peuvent être résumées comme suit :

- la sémantique d'une phrase respecte l'ordre dans lequel les mots sont émis contrairement à une tâche de traduction où les mots traduits peuvent avoir un ordre complètement différent de l'ordre des mots de la phrase source selon la langue.
- dans une tâche de traduction, un mot source peut être aligné à aucun mot cible (fertilité = 0), alors que pour la compréhension chaque mot doit être aligné à un concept, sachant que les mots qui ne contribuent pas au sens de la phrase sont étiquetés par un concept spécifique "NULL".
- dans une tâche de traduction le traitement des mots inconnus (mots hors vocabulaire) est simplifié par le glissement du mot vers la sortie (ce qui est souvent une option favorable, par exemple dans le cas des noms propres ou des entités nommées en général). Pour la compréhension il est impératif de proposer une hypothèse de concept même en présence de mots inconnus. Et pour cela la capacité a prendre en compte le contexte du mot, mais aussi les hypothèses précédentes, est bien souvent un avantage.
- enfin, la mesure d'évaluation est différente entre les deux tâches (BLEU pour la traduction vs. CER pour la compréhension) et donc les outils utilisés pour l'optimisation des systèmes de traduction sont adaptés pour optimiser le score BLEU. Bien sur le recours au TER (moins répandu comme mesure de performance des systèmes de traduction) permet d'utiliser des mesures comparables.

Afin d'améliorer la performance de l'approche à base de segments (PB-SMT) pour la compréhension, nous proposons un certain nombre d'adaptations pour prendre en compte les caractéristiques de la tâche.

Le modèle de réordonnancement fait partie des composants de l'approche PB-SMT. Ce module permet de réordonnancer les phrases source en respectant l'ordre dans lequel leurs traductions en langue cible sont censées apparaitre. Cette étape, indispensable pour une tâche de traduction, peut être pénalisante dans le cas de l'utilisation des PB-SMT pour la compréhension où aucun réordonnancement n'est requis.

En suivant l'hypothèse que la sémantique d'une phrase respecte l'ordre dans lequel les mots sont émis, nous proposons de mettre une contrainte de monotonie pendant

la traduction (décodage monotone). Cette contrainte oblige le décodeur à respecter, en fonction de l'ordre des mots initiaux, l'ordre des concepts générés.

Une difficulté majeure du processus de traduction automatique est l'alignement d'un mot de la langue source avec le mot correspondant dans le langue cible. Vu que les corpus utilisés pour apprendre des systèmes de traduction sont des corpus alignés au niveau des phrases, une étape d'alignement automatique est nécessaire pour obtenir l'alignement en mots. Cependant, la plupart des corpus de compréhension sont étiquetés (alignés) au niveau des segments conceptuels et donc l'utilisation de ces informations d'alignement peut être avantageuse pour aider le processus d'alignement.

Pour cela nous proposons d'utiliser les corpus en format BIO (décrit dans la section 2.3). Ce format garanti que chaque mot de la phrase source soit aligné à son concept correspondant et donc aucun alignement automatique supplémentaire n'est désormais requis. De cette façon, l'extraction de la table de segments est obtenue par un corpus avec un alignement parfait (non bruité).

Un autre problème se pose au niveau de l'optimisation des paramètres du modèle log-linéaire dans l'approche PB-SMT. MERT étant conçu pour une tâche de traduction, il cherche à optimiser les paramètres du modèle en optimisant le score BLEU.

Vu que nous cherchons à évaluer les hypothèses générées par cette approche du point de vue de la compréhension (la mesure d'évaluation du système de compréhension étant le CER et non pas le score BLEU) nous proposons de modifier l'optimisation MERT pour optimiser le CER directement.

Il est important de mentionner qu'une différence majeure entre les deux approches est dans le traitement de mots hors vocabulaires. Dans PB-SMT, les mots hors vocabulaire rencontrés durant le décodage sont projetés "sans traduction" dans la sortie. Pour une tâche de traduction cette projection est avantageuse dans les cas où ces mots hors vocabulaire sont des entités nommées comme nom de personnes, de lieux, etc. Récupérer ces entités en sortie peut former une bonne traduction du fait que leur traduction est la même dans plusieurs langues (quand il s'agit de traduction entre des langues latines par exemple).

En compréhension, le vocabulaire de sorties possibles est limité aux concepts prédéfinis et donc la projection des entités nommées en sortie ne peut en aucun cas être avantageuse. D'autre part, il est possible d'anticiper le problème en enrichissant le corpus d'apprentissage par des listes d'entité nommées relatives au domaine.

Par exemple, dans le contexte de réservation d'hôtel, un nombre important de mots hors vocabulaire viennent des noms de villes absentes dans les données d'apprentissage. Donc nous proposons de rajouter une liste de villes aux données d'apprentissage du système SLU/PB-SMT.

6.2.2 Application à la portabilité multilingue

Dans le chapitre 4 nous avons proposé des méthodes pour porter l'annotation sémantique d'un corpus existant vers un corpus traduit. Cette annotation repose sur des mots afin de pouvoir apprendre des modèles CRFs. L'utilisation de l'approche PB-SMT pour la compréhension peut être aussi utilisée pour la portabilité multilingue des systèmes de compréhension, de façon complémentaire par rapport aux méthodes proposées auparavant.

En effet, cette approche a l'avantage de ne pas recourir à un alignement au niveau des mots, et donc pour l'appliquer dans la méthode TrainOnTarget (voir la section 4.3.2), il suffit de traduire les phrases de la langue source vers la langue cible et ensuite d'apprendre un traducteur sur le corpus parallèle "langue cible | | concept". Cela peut être rapproché de la méthode proposée dans la section 4.4.1. Autrement dit cette technique s'affranchit du besoin de porter l'annotation au niveau mot de la source vers la cible, puisque seule une annotation au niveau de la phrase est nécessaire.

6.3 Méthode de compréhension pour la traduction

Dans cette approche, le problème de la traduction est considéré comme un problème d'étiquetage de la phrase source, sauf que les étiquettes possibles sont tous les mots de la langue cible.

L'apprentissage d'un étiqueteur fondé sur une approche CRF pour une tâche de traduction nécessite un corpus annoté (traduit) au niveau des mots. L'application des modèles IBM nous permet d'obtenir des alignements en mots à partir d'un corpus bilingue aligné au niveau des phrases. Comme pour la compréhension, où plusieurs mots peuvent être associés à un seul concept, plusieurs mots source peuvent être alignés avec un seul mot cible. Pour gérer cela, la proposition la plus simple est d'appliquer la même méthode utilisée pour la compréhension : le passage au format BIO (voir section 2.3). Ainsi la séquence française "je voudrais" qui est alignée au mot italien "vorrei" sera représentée comme : <je, B_vorrei> <voudrais, I_vorrei>.

La difficulté principale pour apprendre des modèles CRFs pour la traduction est liée au nombre élevé d'étiquettes (correspondant à la taille du vocabulaire de la langue cible). (Riedmiller and Braun, 1993) ont proposé d'utiliser l'algorithme RPROP pour l'optimisation des paramètres de modèles lorsqu'il s'agit d'un modèle avec un nombre important de paramètres. Cet algorithme réduit le besoin en mémoire par rapport à d'autres algorithmes d'optimisation (Turian et al., 2006).

Un autre défaut important de l'utilisation des CRFs pour la traduction, est le fait que ce modèle représente un modèle de traduction, mais ne peut pas prendre en compte le réordonnancement des mots ni le modèle de langage cible. Afin d'obtenir un système de traduction efficace à base de CRFs, (Lavergne et al., 2011) ont proposé un modèle CRFs fondé sur des transducteurs à états finis qui composent les différents composants

du système de traduction. Une description de ce modèle est présentée dans la section suivante.

6.3.1 Modèle du LIMSI (FST/CRF)

Les modèles de traduction à base de n-grammes (Mariòo et al., 2006) sont des alternatives aux modèles de traduction à base de segments qui diffèrent dans la manière de représenter le modèle de traduction. Dans cette approche, le modèle de traduction est fondé sur des unités bilingues nommées "tuple". Les tuples sont des paires de séquences de mots où l'une est la traduction de l'autre. Une version discriminante (à base de CRFs) de l'approche n-grammes à été proposé par le LIMSI (Lavergne et al., 2011) pour modéliser $P(t \mid s)$ au lieu de $P(t, s)$.

Le décodeur proposé pour ce modèle est une composition de transducteurs à états finis (Weighted Finite State Transducer, WFST) à l'aide des fonctionnalités standards disponibles dans des bibliothèques comme OpenFST (Allauzen et al., 2007). Ces bibliothèques implémentent les opérations de base sur les WFSTs, comme la projection gauche (π_1), la projection droite (π_2) ou encore la composition ($\circ$).

L'opération de projection d'un transducteur permet d'obtenir un accepteur avec les transitions étiquetées comme les entrées du transducteur dans le cas de projection gauche ou comme les sorties dans le cas de projection droite.

La composition de deux transducteurs consiste à aligner les symboles de sortie du premier transducteur avec les symboles d'entrée du second et de produire un transducteur qui prend en entrée les symboles du premier et donne en sortie les symboles du second. Plus d'informations sur ces opérations et la théorie des WFST en général peuvent être trouvées dans (Mohri, 2009).

Essentiellement, le décodeur de traduction est une composition de transducteurs qui représentent les étapes suivantes :

1. le réordonnancement et la segmentation de la phrase source
2. l'application du modèle de traduction avec une valuation des hypothèses à base de CRFs
3. la composition avec un modèle de langage cible.

(Kumar and Byrne, 2003) ont proposé une architecture assez similaire qui utilise un modèle ATTM (Alignment Template Translation Models) au lieu des CRFs comme modèle de traduction. Cette architecture permet de voir la traduction d'une phrase comme une composition de transducteurs dans l'ordre suivant :

$$\lambda_{traduction} = \lambda_S \circ \lambda_R \circ \lambda_T \circ \lambda_F \circ \lambda_L$$

sachant que

- λ_S est l'accepteur de la phrase source s ;
- λ_R implémente les règles de segmentation et de réordonnancement ;
- λ_T est un dictionnaire de tuple, qui associe des séquences de la langue source avec leurs traductions possibles en se basant sur l'inventaire des tuples ;
- λ_F est une fonction d'extraction de motifs (feature matcher), qui permet d'attribuer des scores de probabilité aux tuples en les comparant aux motifs du modèle CRF ;
- λ_L est un modèle de langage de la langue cible.

Nous allons passer en revue chacun de ces composants afin de spécifier son rôle.

Le modèle de réordonnancement λ_R

Le modèle de réordonnancement représente les règles de réécriture de la phrase. Le modèle λ_R est appris suivant l'approche proposée par (Crego and Mariño, 2006). Cette approche propose de se fonder sur l'étiquetage grammatical de la phrase source (Part-Of-speech, POS) pour réordonnancer la partie source du corpus d'apprentissage.

Dans cette approche, le modèle de réordonnancement représente les règles de réécriture de la phrase $t_1, ..., t_n$ –> $i_1, ..., i_n$, sachant que $t_1, ..., t_n$ est une séquence d'étiquettes grammaticales attribuées aux mots de la phrase source et $i_1, ..., i_n$ est une séquence de positions dans laquelle la phrase source sera réordonnancée.

Chaque cas de réordonnancement observé dans le corpus d'apprentissage est généralisé comme une règle non-déterministe qui représente un réordonnancement possible d'une sous-partie de la phrase source. La figure 6.1 montre comment l'on peut extraire des règles de réordonnancement à partir de phrases parallèles. L'extraction d'un règle repose sur l'étiquetage grammatical de la phrase source et l'alignement mot-à-mot entre la phrase source et la phrase cible. Dans cet exemple le mot "intéressante" dans la position "2" est aligné au mot "interesting" dans la position "1" et donc la règle DT NS AC –> 0 2 1 peut être extraite.

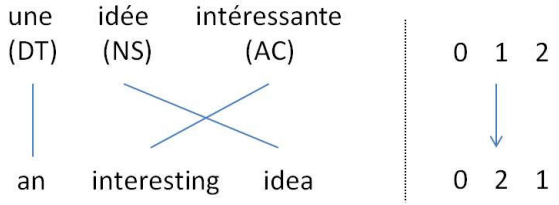

FIGURE 6.1 – *Exemple d'extraction de règle de réordonnancement*

Chaque règle est représentée par un transducteur à états finis et l'ensemble des réordonnancements possibles de mots est considéré comme la composition de ces transducteurs. Au final, λ_R est obtenu par la composition des différentes règles de réordonnancement avec le transducteur qui représente toutes les segmentations possibles de la phrase d'entrée.

La figure 6.2 montre un exemple de graphe de recherche composé. La première ligne de cette figure montre la phrase d'entrée. Dans la deuxième ligne le graphe est étendu par la règle "NS AC –> 1 0" et enfin, la troisième ligne montre l'extension du graphe obtenue par la règle "NS AC CC AC –> 1 2 3 0". Ainsi la phrase présentée dans cet exemple peut avoir trois versions de réordonnancement :

1. idée claire et intéressante
2. claire idée et intéressante
3. claire et intéressante idée

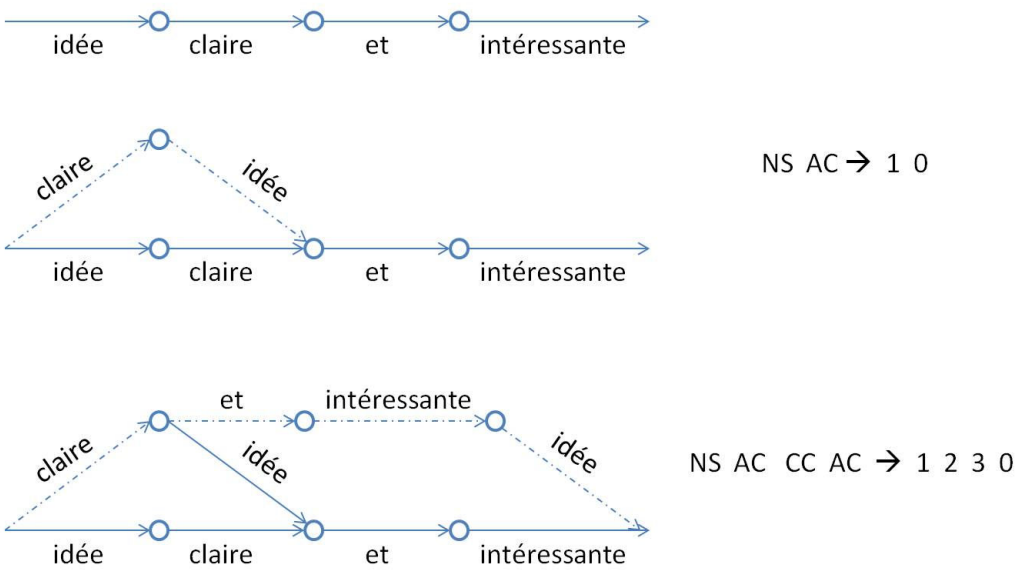

FIGURE 6.2 – *Exemple de graphe composé de plusieurs chemins de réordonnancement*

La sortie de $\lambda_S \circ \lambda_R$ est une séquence de tuples sources $\hat{s}$. Chaque chemin de ce transducteur est pondéré par un modèle de segmentation estimé sur la partie source du corpus d'apprentissage. Il est important de noter que ces scores sont normalisés, ainsi le poids de chaque chemin $\hat{s}$ de la composition $\lambda_S \circ \lambda_R$ correspond à la probabilité $logP(\hat{s} \mid s)$.

L'extraction des tuples λ_T :

Les tuples sont extraits à partir des phrases parallèles en deux étapes. D'abord une procédure itérative permet de regrouper les mots. Chaque groupe est composé des mots de la phrase source qui sont alignés à un seul mot ou un groupe de mots dans la phrase cible. Cette procédure itère jusqu'au moment où aucun nouveau groupe ne peut être obtenu. Dans un second temps les tuples sont extraits à partir de ces groupes en respectant l'ordre des mots dans la phrase cible. Donc l'extraction de tuple se fait à partir de paires de phrases alignées mot-à-mot en respectant les conditions suivantes :

- la segmentation des phrases est monotone ;
- aucun mot du tuple n'est aligné à un mot en dehors du tuple ;
- aucun tuple plus court ne peut être extrait en respectant les conditions précédentes.

Contrairement à la segmentation en "phrases" dans la traduction à base de segment, les conditions précédentes rendent la segmentation d'une paire de phrase unique (Khalilov and Fonollosa, 2009).

La fonction d'extraction de motifs (feature matcher) λ_F :

λ_F est le feature matcher. Il permet d'attribuer des scores de probabilité au tuple en se basant sur les fonctions caractéristiques d'un modèle CRFs. Ce transducteur est représenté par une série de transducteurs pondérés, chaque transducteur étant responsable d'une classe donnée de fonctions (les fonctions uni-grammes et bi-grammes des CRFs). Le plus simple transducteur de cette famille est celui qui gère la classe des fonctions uni-grammes, c'est-à-dire les fonctions qui correspondent uniquement aux observations et aux étiquettes courantes. Les dépendances entre les symboles source et/ou cible peuvent être représentées par des transducteurs à états finis tels que ceux montrés dans la figure 6.3, qui représente respectivement les fonctions bigrammes de la cible et les fonctions bigrammes conjointes entre la source et la cible.

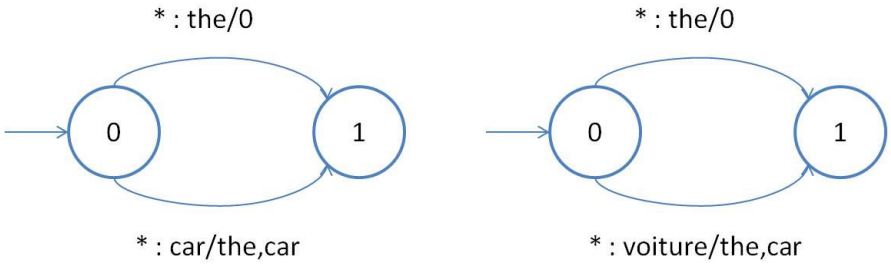

FIGURE 6.3 – *Exemple de feature matcher. Le symbole (*) représente toute observation possible.*

Le feature matcher λ_F est calculé comme une composition de ces transducteurs élémentaires, où seules les étiquettes sources et cibles qui peuvent apparaitre dans une phrase d'entrée courante sont considérées.

En suivant les notations utilisées dans (Mohri et al., 2002), les transducteurs sont définis par la structure algébrique de semi-anneau. Un semi-anneau (semiring) est constitué d'un ensemble K avec une opération associative et commutative $\oplus$ et une opération associative $\otimes$ ainsi que deux éléments neutres $\overline{0}$ et $\overline{1}$.

Les poids du modèle λ_F sont interprétés par un semi-anneau tropical (tropical semiring) et exp λ_F est obtenu par le remplacement des poids w dans λ_F par $exp(w)$ dans le semi-anneau réel (real semiring).

Le décodage d'un graphe

La segmentation et le réordonnancement de la source produisent un graphe avec de multiples chemins . Si la sortie de $(\lambda_S \circ \lambda_R \circ \lambda_T)$ était unique, la traduction de la phrase source serait le meilleur chemin de $\lambda_S \circ \lambda_R \circ \lambda_T \circ exp(\lambda_F)$. Dans le cas d'une application réelle avec des multiples segmentations possibles nous cherchons :

$$\text{argmax}_{\hat{t}} P(\hat{t} \mid s) = \text{argmax}_{\hat{t}} \sum P(\hat{t}, \hat{s} \mid s)$$

$$\text{argmax}_{\hat{t}} P(\hat{t} \mid s) = \text{argmax}_{\hat{t}} \sum P(\hat{t} \mid \hat{s}) P(\hat{s} \mid s)$$

Cela exige de comparer les scores entre les différentes segmentations de la source $\hat{s}$ et donc de calculer $P(\hat{t} \mid \hat{s})$ et $P(\hat{s} \mid s)$. La valeur de normalisation Z pour toutes les séquences de $\lambda_S \circ \lambda_R$ peut être obtenue par des opérations de base sur les transducteurs à états finis et donc peut être définie comme :

$$\lambda_D = det(\pi_1(\pi_2(\lambda_S \circ \lambda_R) \circ \lambda_T \circ exp(\lambda_F)))$$

L'opération de détermination permet dans ce cas l'accumulation des valeurs Z correspondant à chaque chemin $\hat{s}$.

En remplaçant chaque poids w par un $-log(w)$ dans λ_D, et en utilisant un semi-anneau logarithmique nous pouvons calculer le $-log(\lambda_D)$. Au final le meilleur chemin peut être obtenu par :

$$BestPath(\pi_2(\lambda_S \circ \lambda_R) \circ -log(\lambda_D) \circ \lambda_T \circ \lambda_F)$$

Décodage avec un modèle de langage cible

Une manière de prendre en compte le modèle de langage cible pendant le décodage est de composer le graphe de traduction présenté précédemment avec un graphe qui représente le modèle de langage. Dans ce cas, le décodage peut être présenté comme :

$$BestPath(\pi_2(\lambda_S \circ \lambda_R) \circ -log(\lambda_D) \circ \lambda_T \circ \lambda_F \circ \lambda_L)$$

6.4 Décodage conjoint pour la traduction et la compréhension, cas de la méthode de portabilité TestOnSource

Notre étude des relations entre les différentes approches est réalisée avec l'objectif de pouvoir les combiner du mieux possible pour la portabilité multilingue d'un système de compréhension.

Dans les chapitres précédents, nous avons montré que la meilleure méthode pour porter un système de compréhension existant vers une nouvelle langue (nommée auparavant TestOnSource, présentée dans la section 4.3.1) est de traduire les énoncés utilisateurs de la nouvelle langue vers la langue du système existant et ensuite de faire étiqueter les énoncés (traduits) par ce système.

Notre proposition a été basée sur une cascade d'un système de traduction (PB-SMT) et d'un système de compréhension (CRFs). La meilleure hypothèse générée par le système de traduction est mise en entrée du système de compréhension. Cependant, d'autres hypothèses de la liste des n-meilleures traductions peuvent différer uniquement dans l'ordre des mots et cet ordre-là peut être mieux interprété par l'étiqueteur sémantique. Donc cette sélection a priori de la meilleure traduction n'optimise pas forcément le comportement du système global.

Pour faire face à ce problème nous proposons d'effectuer un décodage conjoint entre la traduction et la compréhension d'une manière comparable à la méthode TestOnSource. Ce décodage conjoint aura l'avantage de pouvoir optimiser la sélection de traduction en prenant compte des étiquettes qui peuvent être attribuées aux différentes traductions possibles, alors que dans TestOnSource cela était fait en deux étapes séparées.

Ce problème rejoint, dans son esprit, le problème de cascade des composants d'un système de dialogue. Dans une architecture standard, le système de reconnaissance de la parole transmet sa meilleure hypothèse de transcription au système de compréhension. Vu que cette hypothèse est bruitée, elle n'est pas forcément l'hypothèse que le système de compréhension pourra étiqueter le mieux.

Plusieurs travaux ont proposé un décodage conjoint entre la reconnaissance et la compréhension de la parole pour prendre en compte des n-meilleures hypothèses de reconnaissance lors de l'étiquetage sémantique. Ces premiers travaux (Tür et al., 2002;

Servan et al., 2006; Hakkani-Tür et al., 2006) ont proposé d'utiliser un réseau de confusion entre les différentes sorties de reconnaissance pour obtenir un graphe d'hypothèses. Le système de compréhension dans ces propositions a été représenté par un FST qui transforme des mots en concepts (voir 2.4.2.3). Un décodage conjoint peut être obtenu par la composition du graphe de reconnaissance avec le graphe de compréhension.

Les résultats positifs obtenus par ces propositions, ont encouragé d'autres travaux dans la même ligne. Vu que les modèles les plus performants dans la littérature sont des modèles discriminants (contrairement au FST, de nature générative), (Anoop Deoras and Hakkani-Tur, 2012) ont proposé d'utiliser des modèles CRFs au lieu des FST pour la compréhension.

Dans la lignée de ces travaux, nous cherchons à obtenir un décodage conjoint pour la traduction et la compréhension. Les deux systèmes étant de natures différentes, leur combinaison et leur optimisation conjointe sont rendues délicates, d'où l'intérêt d'uniformiser les systèmes pour les deux tâches.

La proposition d'utiliser une approche FST/CRF pour la traduction peut être étendue pour la compréhension. Donc un système de compréhension $\lambda_{comprehension}$ peut être obtenu par une approche FST/CRF de la même manière que proposé dans la section 6.3.1 et en suivant les adaptations entrevues dans la section 6.2.1. Cette représentation nous permet d'obtenir un graphe de compréhension similaire à celui obtenu pour la traduction.

Vu que le vocabulaire des sorties du graphe de traduction est le même que celui de l'entrée du graphe de compréhension, ces deux graphes peuvent être composés facilement en utilisant la fonction de composition ($\circ$) pour donner un graphe permettant de décodage conjoint :

$$\lambda_{conjoint} = \lambda_{tradcution} \circ \lambda_{comprehension}$$

Cela peut être détaillé comme :

$$\lambda_{conjoint} = (\lambda_S \circ \lambda_R \circ \lambda_T \circ \lambda_F \circ \lambda_L)_{traduction} \circ (\lambda_S \circ \lambda_R \circ \lambda_T \circ \lambda_F \circ \lambda_L)_{comprehension}$$

Cette composition est comparable à la méthode TestOnSource : elle prend une phrase de la langue cible en entrée et attribue une séquence de concept à cette phrase en passant par un étiqueteur disponible dans la langue source. Elle nous permet d'obtenir un décodage conjoint entre la traduction et la compréhension du moment où les probabilités des deux graphes sont prises en compte. Un tel décodage ne cherche pas à optimiser la traduction en soi, mais à optimiser le choix d'une traduction qui donnera une meilleure compréhension.

A la lecture de l'équation de $\lambda_{conjoint}$, nous remarquons que l'accepteur $\lambda_{S\,comprehension}$ du graphe de compréhension prend en entrée un graphe de traduction contrairement à l'accepteur de traduction $\lambda_{S_{traduction}}$ qui prend une chaine en entrée.

	SMT	**SLU**
modèle	PB-SMT/log-linéaire	CRFs
alignement	segments (phrases)	mots
réordonnancement	oui	non
modèle de langage sur les sorties	5-grammes	bi-grammes (intégré au décodage)
mots HV	projection	traitement
paramètres	14	centaine de milliers
ajustement des paramètres	MERT (heuristiques)	RPROP (gradient)

TABLE *6.1* – *Comparaison entre les approches utilisées en traduction (SMT) et en compréhension de la parole (SLU).*

Le $\lambda_{conjoint}$ peut être généralisé pour composer un graphe de reconnaissance de la parole avec un graphe de compréhension dans le cadre d'un système de dialogue. Dans ce cas des procédures d'élagage devront être prises en compte afin d'assurer que les opérations de composition puissent être réalisées selon les contraintes classiques (temps de calcul et espace mémoire machine disponible).

En plus du fait qu'il nous permet d'obtenir un décodage conjoint, le passage des CRFs en graphe FST/CRF permet d'aboutir à un modèle CRFs scorant l'espace de recherche complet, ce qui est extrêmement complexe avec les implémentations CRFs classiques.

6.5 Conclusion

Dans ce chapitre nous avons présenté une comparaison théorique entre une approche générative et une autre discriminante pour deux tâches : la traduction automatique et la compréhension de la parole. Un modèle CRFs a été comparé à un modèle log-linéaire pour réaliser ces tâches.

La compréhension de la parole peut être vue comme une traduction d'une chaine de mots vers une chaîne de concepts et la traduction automatique peut être vue comme un étiquetage sémantique des phrases de la langue source par des étiquettes qui correspondent au vocabulaire de la langue cible. Malgré cette représentation, chaque tâche a des spécificités qui doivent être prises en compte afin de pouvoir optimiser ses performances.

Les caractéristiques de ces tâches ont été analysées dans ce chapitre et une comparaison entre ces deux tâches peut être récapitulée dans le tableau 6.1.

Nous avons proposé plusieurs adaptations sur le modèle PB-SMT afin de pouvoir prendre en compte les caractéristiques de la compréhension (alignement, réordonnancement, ajustement de poids et traitement de mots hors vocabulaire). D'autre part nous avons proposé d'utiliser une approche à base de FST afin de pouvoir utiliser un modèle CRFs pour la traduction. Le modèle FST/CRF permet de composer le modèle

de traduction avec un modèle de réordonnancement et un modèle de langage afin de se comparer au modèle PB-SMT.

Dans le but de vouloir obtenir un décodage conjoint entre la traduction et la compréhension (dans le cas de notre scénario lié à la portabilité selon la méthode TestOn-Source), nous avons proposé d'uniformiser les approches utilisées pour les deux modèles en utilisant l'approche FST/CRF pour les deux tâches. Un décodage conjoint peut être obtenu par la composition des deux graphes.

Une évaluation des approches proposées dans ce chapitre est présentée dans le chapitre 7.

Chapitre 7

Approches conjointes : expériences et résultats

Sommaire

7.1 Introduction . . . 118
7.2 Evaluation des systèmes de traduction à base de segments pour une tâche de compréhension . . . 118
7.2.1 Evaluation du système français . . . 118
7.2.2 Evaluation du système italien . . . 120
7.3 Evaluation des systèmes de traduction à base de CRFs . . . 121
7.3.1 Evaluation du système de traduction français vers italien . . . 122
7.3.2 Evaluation du système de traduction italien vers français . . . 123
7.4 Evaluation des systèmes de traduction selon l'approche FST/CRF . . . 124
7.4.1 Evaluation pour une tâche de traduction . . . 124
7.4.2 Evaluation pour une tâche de compréhension . . . 126
7.5 Décodage conjoint dans le cas d'un scénario de portabilité du français vers l'italien d'un système de compréhension (TestOnSource) 126
7.6 Conclusion . . . 129

7.1 Introduction

Afin de pouvoir comparer les approches génératives et les approches discriminantes sur des tâches similaires, nous proposons de reprendre le même protocole expérimental adopté dans le chapitre 5. Nos expériences seront basées sur le corpus MEDIA français et la partie traduite de ce corpus vers l'italien.

Dans un premier temps, nous proposons d'utiliser un modèle de traduction à base de segments pour apprendre des systèmes de compréhension. Ces systèmes seront comparés aux systèmes de référence présentés dans la section 5.3. L'évaluation de ces systèmes sera présentée dans la section 7.2.

Ce même corpus avec sa traduction italienne sera utilisé comme corpus parallèle pour apprendre des systèmes de traduction à base de CRFs. La comparaison de ces systèmes avec les systèmes de traduction à base de segments sera présentée dans la section 7.3.

L'utilisation d'une approche FST/CRF permet d'améliorer les performances du système de traduction par rapport à une approche CRFs de base. Cette approche peut aussi être utilisée pour une tâche de compréhension, en gérant des graphes d'hypothèse en entrée. La mise en place de ce modèle, ainsi que son évaluation pour une tâche de traduction et pour une tâche de compréhension seront présentés dans la section 7.4.

Enfin, notre proposition pour un décodage conjoint pour la traduction et la compréhension est évaluée dans la section 7.5.

7.2 Evaluation des systèmes de traduction à base de segments pour une tâche de compréhension

Afin d'évaluer l'approche PB-SMT pour une tâche de compréhension, nous proposons d'apprendre deux systèmes différents. Le premier pour le français (appris sur la totalité du corpus MEDIA) et le second pour l'italien (appris sur le corpus MEDIA traduit en italien). Nous avons aussi appliqué nos propositions d'améliorations progressives pour cette approche sur les deux systèmes.

Nous avons enfin proposé de combiner les sorties du système italien avec les sorties des autres approches de portabilité présentées dans la section 4.3 pour avoir des hypothèses complémentaires dans le réseau de confusion décrit dans 5.3.5.

7.2.1 Evaluation du système français

Nos premières tentatives pour construire le modèle PB-SMT pour la compréhension du français ont clairement montré des performances inférieures aux CRFs (CER=23,2% après réglage des paramètres MERT pour le PB-SMT à comparer aux 12,9% pour les

Modèle	Sub	Del	Ins	CER
Initial	5,4	4,1	14,6	24,1
+MERT (BLEU)	5,6	8,4	9,2	23,2
+Décodage monotone	6,2	7,8	8,7	22,7
+Format BIO	5,7	9,3	5,3	20,3
MERT (CER)	5,3	9,2	4,6	19,1
Traitement de mots HV	5,8	7,4	5,1	18,3

TABLE 7.1 – *Les améliorations itératives du modèle SLU/PB-SMT sur le corpus* MEDIA *français(CER%).*

Modèle	Sub	Del	Ins	CER
CRFs	3,1	8,1	1,8	12,9
PB-SMT	5,8	7,4	5,1	18,3

TABLE 7.2 – *Comparaison des types d'erreurs entre l'approche CRFs et l'approche PB-SMT (CER%).*

CRFs). Les améliorations progressives du modèle proposées dans la section 6.2 sont évaluées dans le tableau 7.1.

L'utilisation de la contrainte de monotonie durant le décodage permet une réduction de 0,5% absolu. Convertir les données selon le formalisme BIO avant la phase d'apprentissage réduit le CER de façon significative de 2,4%. Enfin, optimiser le score CER à la place du score BLEU réduit le CER de 0,4% supplémentaire. Enfin, l'ajout d'une liste de villes à l'ensemble d'apprentissage avant réapprentissage du modèle PB-SMT permet une réduction finale de 0,8%.

Les résultats montrent qu'en dépit de réglages fins de l'approche PB-SMT, les approches à base de CRFs obtiennent toujours les meilleures performances pour une tâche de compréhension (12,9% de CER pour CRFs vs. 18,3% de CER pour PB-SMT).

A partir d'une analyse rapide du type d'erreur de chaque modèle, nous pouvons observer (voir tableau 7.2) que les méthodes utilisant des CRFs ont un haut niveau de suppressions comparativement aux autres types d'erreurs, tandis que la méthode PB-SMT présente un meilleur compromis entre les erreurs de suppression et d'insertion, et ce bien qu'elle aboutisse à un CER plus élevé.

Un nombre important d'erreurs causées par le modèle PB-SMT pour la compréhension et dû à une mauvaise segmentation (souvent sur-segmentation) des phrases. Par exemple une phrase comme "Je voudrais réserver en fait pour la ville de Nice du premier aux trois novembre" a une insertion venant d'une sur-segmentation de la phrase. Il est important de mentionner que cette phrase est étiquetée parfaitement par un modèle CRFs (voir la figure 7.1).

Cette caractéristique des modèles SLU/PB-SMT mène à une distribution équilibrée d'erreurs entre les omissions, les insertions et les substitutions, alors que pour les CRFs

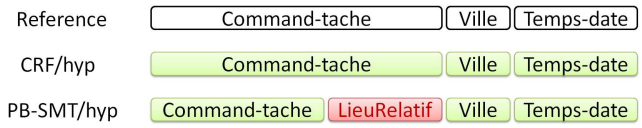

FIGURE 7.1 – *Comparaison entre l'hypothèse générée par le modèles SLU/CRF et celle générée par le modèle SLU/PB-SMT pour la phrase "Je voudrais réserver en fait pour la ville de Nice du premier aux trois novembre".*

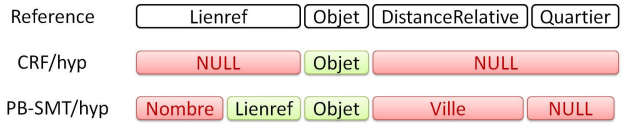

FIGURE 7.2 – *Comparaison entre l'hypothèse générée par le modèles SLU/CRF et celle générée par le modèle SLU/PB-SMT pour la phrase "un autre quartier du côté de Saint-Michel".*

le plus grand nombre d'erreur venait des omissions. Par exemple la phrase "un autre quartier du côté de Saint-Michel" génére trois omissions par un modèle CRFs, tandis qu'elle génére une insertion, une subtitution et une omission par un modèle PB-SMT (voir figure 7.2).

7.2.2 Evaluation du système italien

Un résultat très comparable est obtenu par l'application de l'approche PB-SMT pour la compréhension de l'italien. Le modèle PB-SMT, après réglage des paramètres par l'algorithme MERT, donne un CER de 28,1% (à comparer à 19,9% pour les CRFs). Les améliorations progressives du modèle (proposées en section 6.2) sont évaluées dans le tableau 7.3.

L'utilisation de la contrainte de monotonie durant l'alignement en mot permet une réduction de 0,6% absolu. Convertir les données selon le formalisme BIO avant la phase d'apprentissage réduit le CER de façon significative de 2,8%. Enfin optimiser le CER à la place du BLEU réduit le CER de 0,3% supplémentaire. En terme de traitement des mots hors vocabulaire, l'ajout d'une liste de villes à l'ensemble d'apprentissage avant réapprentissage du modèle PB-SMT permet une réduction finale de 0,5%.

Comme pour le français, les approches à base de CRFs obtiennent toujours les meilleures performances. Vu que le type d'erreur n'est pas le même pour les modèles PB-SMT et les modèles CRFs, nous pensons que la combinaison des deux peut augmenter la performance globale obtenue. Cette combinaison sera réalisée avec les différentes méthodes proposées pour la portabilité du système de la compréhension.

Pour cela nous proposons de combiner les deux approches principales (TestOnTarget et TrainOnTarget) avec l'approche PB-SMT afin de bénéficier de leurs caractéristiques respectives pour améliorer la performance globale.

Modèle	Sub	Del	Ins	CER
Initial	6,5	4,0	18,6	29,1
+MERT (BLEU)	6,3	9,3	12,5	28,1
+Décodage monotone	7,4	8,4	11,8	27,5
+Format BIO	6,5	10,6	7,7	24,7
MERT (CER)	6,4	10,9	7,2	24,4
Traitement de mots HV	7,2	10,5	6,1	23,9

TABLE 7.3 – *Les améliorations itératives du modèle SLU/PB-SMT sur le corpus* MEDIA *italien(CER%).*

Modèle	Sub	Del	Ins	CER
BASIQUE	6,2	9,7	2,7	18,6
TOUT	5,4	10,5	2,3	18,2
TOUT-SLU/PB-SMT	6,6	10,2	2,3	19,1

TABLE 7.4 – *Combinaison des système de compréhension de l'italien, avec et sans l'approche PB-SMT.*

La combinaison (dénotée BASIQUE dans le tableau 7.4) est simple : un réseau de confusion est construit à partir des trois hypothèses et la séquence de concepts correspondant à la plus grande probabilité a posteriori est calculée. La performance est améliorée de façon significative (-1,3% CER) ce qui confirme la complémentarité des méthodes.

Finalement nous combinons toutes les méthodes proposées dans ce papier (TrainOnTarget, TestOnSource, +SCTD, +SPE, +SCTD+SPE, PB-SMT). Ce qui permet d'atteindre les meilleures performances rapportées sur le test MEDIA (18,2% vs 19,3% reporté pour une approche CRFs robuste dans le tableau 5.9).

Afin de mesurer l'influence de la méthode PB-SMT sur les performances de la combinaison, nous avons aussi évalué les performances de la combinaison pour laquelle l'approche PB-SMT n'est pas utilisée. Cette expérience montre qu'en dépit de résultats individuels moyens, la méthode PB-SMT a une influence importante sur la combinaison.

7.3 Evaluation des systèmes de traduction à base de CRFs

Notre proposition d'utiliser une approche CRF pour la traduction a été évaluée sur deux systèmes : le premier français vers italien et le second italien vers français.

Pour cela nous utilisons la partie traduite manuellement (du français vers l'italien) du corpus MEDIA comme corpus parallèle pour apprendre le modèle de traduction. L'outil GIZA++ (Och, 2003a) a été utilisé pour apprendre un alignement mots à mots

Modèle	Feature	BLEU
PB-SMT	-	43,62
CRF/uni-grammes	U	37,76
CRF/bi-grammes	U + B	39,62

TABLE 7.5 – *Evaluation du modèle CRFs pour la traduction du français vers l'italien (BLEU %).*

entre les deux corpus et l'outil Wapiti (Lavergne et al., 2010) a été utilisé pour apprendre les modèles CRFs.

Nous effectuons une comparaison entre ces modèles et des modèles PB-SMT dans des conditions similaires. L'ensemble de test de MEDIA français avec sa traduction manuelle en italien est utilisé pour évaluer les modèles.

7.3.1 Evaluation du système de traduction français vers italien

Dans un premier temps, nous cherchons à apprendre un modèle CRFs pour la traduction, en utilisant l'algorithme de RPROP comme proposé dans la section 6.3. Les performances obtenues par ces modèles sont présenté dans le tableau 7.5. Comme escompté, les résultats montrent que le modèle CRF/bi-grammes est plus performant que le modèle CRF/uni-grammes (39,62% vs. 37,76%) mais cette performance reste significativement moins bonne que la performance obtenue par la méthode PB-SMT classique.

Afin d'avoir une comparaison juste entre les deux méthodes, nous cherchons à évaluer l'approche PB-SMT dans les mêmes conditions que l'approche CRFs. Il faut noter que la méthode PB-SMT utilise un modèle de réordonnancement alors que les CRFs, plus dédiés à des étiquetages séquentiels, ne comprennent pas un tel modèle. Pour cela nous rajoutons une contrainte de monotonie dans le décodage pour l'approche PB-SMT empêchant tout réordonnancement.

Il est aussi important de mentionner que l'approche PB-SMT utilise un modèle de langage pour sélectionner la meilleure traduction. Les performances du modèle PB-SMT de référence sont obtenues en utilisant un modèle de langage tri-grammes (utilisé généralement dans les systèmes de traduction). Cependant l'approche CRFs ne permet pas d'utiliser un tel modèle de langage. Les CRFs utilisent deux types de fonction : les uni-grammes et les bi-grammes.

Une fonction uni-grammes permet de prendre en compte une seule étiquette à la fois caractérisant l'association du mot et de l'étiquette. Les fonctions bi-grammes portent sur un couple d'étiquettes successives. Entre les uni-grammes et le bi-grammes le nombre de paramètres du modèle est assez important (de l'ordre de la centaine de milliers), par ailleurs la complexité du décodage est aussi fonction de la complexité portentielle d'association des étiquettes de sorties. Donc les CRFs ne peuvent pas aller au-delà des paramètres de type bi-grammes pour couvrir des tri-grammes.

	CRFs	PB-SMT
-	39,62	43,62
monotone	39,62	42,46
bi-grammes	39,62	42,03
traitement de mots HV	**40,23**	**42,03**

TABLE 7.6 – *Comparaison objective entre le modèle PB-SMT et le modèle CRFs pour la traduction du français vers l'italien (BLEU %).*

Pour évaluer l'approche CRFs et l'approche PB-SMT dans les mêmes conditions, et vu qu'on ne peut pas augmenter la taille des fonctions du modèles CRFs, nous proposons de dégrader l'approche PB-SMT et de réévaluer sa performance en utilisant un modèle de langage de type bi-grammes.

Par ailleurs, en observant les sorties du modèle CRFs, nous remarquons que les mots inconnus (hors-vocabulaire) dans le test ont été traduits par d'autres mots du corpus cible selon le contexte général de la phrase, contrairement à l'approche PB-SMT qui a tendance à projeter les mots hors-vocabulaire tels qu'ils sont dans la phrase traduite. Ces mots hors-vocabulaire, étant dans la plupart des cas des noms de ville ou de lieux, leur traduction ne change pas d'une langue à l'autre, et donc leur projection dans la sortie traduite est avantageuse pour les modèles PB-SMT. Pour cela nous proposons un pré-traitement des mots inconnus dans la phrase source permettant de les récupérer en sortie dans l'approche CRFs. Les résultats des adaptations successives des méthodes PB-SMT et CRFs pour les rendre plus comparables sont présentés dans le tableau 7.6 où la dernière ligne propose une comparaison plus objective des méthodes.

Les résultats présentés dans ce tableau montrent que le décodage monotone dégrade la performance du modèle PB-SMT de 1,16% absolu. L'utilisation d'un modèle de langage bi-grammes augmente la perte de 0,43% supplémentaire. Le traitement des mots hors-vocabulaire permet au modèle CRFs de récupérer 1,61% de score BLEU par rapport au modèle CRFs de référence. On remarque que malgré la dégradation du modèle PB-SMT et les améliorations du modèle CRFs, la performance de ce dernier reste inférieure à celle du modèle PB-SMT (40,23% pour les CRFs vs. 42,03% pour PB-SMT).

7.3.2 Evaluation du système de traduction italien vers français

D'une manière similaire, nous apprenons un modèle CRFs pour la traduction de l'italien vers le français. Afin d'avoir une évaluation comparable à celle présentée pour le système de traduction italien vers français, nous appliquons les mêmes dégradations présentées dans la section précédente sur le modèle PB-SMT (décodage monotone + modèle de langage bi-grammes). Nous appliquons aussi un pré-traitement sur les mots hors-vocabulaire pour les projeter dans la sortie de traduction. Les résultats de ces comparaisons sont présentés dans le tableau 7.7.

Confirmant ce qui a pu être observé précédemment, le modèle PB-SMT reste

	CRFs	PB-SMT
référence	42,53	47,18
monotone	42,53	46,28
bi-grammes	42,53	45,96
traitement de mots HV	**43,47**	**45,96**

TABLE 7.7 – *Comparaison entre le modèle PB-SMT et le modèle CRFs pour la traduction de l'italien vers le français (BLEU %).*

meilleur que le modèle CRFs malgré toutes les contraintes rajoutées sur ce dernier et toutes les améliorations proposées pour le modèle CRFs. Ces résultats encouragent le passage à une approche FST/CRF pour la traduction qui pourra élargir les capacités du modèle CRFs en incluant un modèle de réordonnancement et un modèle de langage plus large.

7.4 Evaluation des systèmes de traduction selon l'approche FST/CRF

Avant d'évaluer notre proposition de décodage conjoint entre la traduction et la compréhension, nous cherchons à évaluer séparément les différents modèles FST/CRF qui seront utilisés dans cette combinaison. Dans un premier temps, nous évaluons cette approche pour une tâche de traduction en apprenant deux systèmes de traduction dans les deux sens français vers italien et italien vers français. Ensuite nous évaluons cette approche pour une tâche de compréhension.

Comme pour les expériences précédentes, le corpus MEDIA avec sa traduction italienne est utilisé pour l'apprentissage et le test. Il est important de mentionner que malgré le fait que nous comparons les performances obtenues par cette nouvelle approche avec les systèmes présentés précédemment, cette comparaison est donnée à titre indicatif car le but de cette approche n'est pas d'avoir les meilleures performances mais d'être dans des bonnes conditions pour pouvoir appliquer un décodage conjoint.

7.4.1 Evaluation pour une tâche de traduction

Un modèle FST/CRF pour la traduction a été construit comme décrit dans la section 6.3.1. Ce modèle a été construit à partir de l'outil n-code (Crego et al., 2011), implémenté pour apprendre des modèles de traduction à base de ngrammes (Mariòo et al., 2006).

Cet outil utilise la bibliothèque OpenFst pour construire un graphe de traduction qui est la composition de 4 transducteurs :

- un accepteur λ_S,
- un modèle de réordonnancement λ_R,
- un modèle de traduction à base de n-grammes λ_T et

Modèle	Langue	BLEU
PB-SMT	FR –> IT	43,62
CRFs		40,23
FST/CRF		41,18
PB-SMT	IT –> FR	47,18
CRFs		43,47
FST/CRF		44,09

TABLE 7.8 – *Comparaison entre les différentes approches (PB-SMT, CRFs, FST/CRF) pour la traduction.*

– un modèle de langage λ_L.

La différence entre le modèle implémenté par cet outil et le modèle FST/CRF qu'on cherche à apprendre est dans les poids du modèle de traduction. Nous adaptons cet outil pour interroger un modèle CRFs afin d'estimer les probabilités de traduction et ensuite nous appliquons une normalisation des scores de probabilité obtenus par ce modèle sur les différents chemins du graphe (comme cela a été proposé dans la section 6.3.1).

Dans n-code le modèle de réordonnancement est appris suivant l'approche proposée par (Crego and Mariño, 2006). Cette approche nécessite un étiquetage grammatical des phrases source et un alignement au niveau des mots entre les phrases source et les phrases cible pour apprendre le modèle λ_R. Nous proposons d'utiliser l'outil TreeTagger (H., 1994) pour obtenir l'étiquetage grammatical et l'outil GIZA++ pour l'alignement en mots.

Le modèle de langage utilisé dans nos expériences et un modèle tri-grammes appris sur la partie cible de notre corpus d'apprentissage à l'aide de l'outil SRILM.

Nous apprenons deux modèles de traduction français vers italien et italien vers français. Le tableau 7.8 présente une comparaison entre trois modèles : le modèle FST/CRF, le modèle PB-SMT (de référence) et le modèle CRFs (présenté dans la section précédente).

Les résultats présentés dans ce tableau montrent que l'approche FST/CRF donne des performances assez comparables à celles obtenues par l'approche PB-SMT. Malgré une dégradation d'environ 2% absolu, ces performances restent assez élevées pour une tâche de traduction (malgré un ensemble d'apprentissage de taille réduite) justifiées dans notre contexte par le vocabulaire limité du domaine.

D'autre part les résultats montrent que l'utilisation de l'approche FST/CRF est doublement avantageuse par rapport à l'utilisation d'une approche CRFs ; en plus du fait que l'approche FST/CRF permet de traiter des graphes, cette approche permet d'augmenter la performance du système d'environ 1% absolu.

Modèle	Sub	Del	Ins	CER
CRFs	3,1	8,1	1,8	12,9
FST/CRF ($\lambda_S \circ \lambda_R \circ \lambda_T \circ \lambda_F \circ \lambda_L$)	4,2	8,8	2,3	15,3
FST/CRF ($\lambda_S \circ \lambda_T \circ \lambda_F$)	3,6	8,2	1,9	13,7
+traitement de mots HV	3,7	7,9	1,8	13,4

TABLE 7.9 – *Evaluation de l'approche FST/CRF pour la compréhension du français.*

7.4.2 Evaluation pour une tâche de compréhension

De la même manière nous apprenons un modèle FST/CRF pour une tâche de compréhension du français. La totalité du corpus MEDIA a été utilisé comme corpus parallèle "mots | | concepts" pour apprendre ce modèle. Un modèle de langage tri-grammes a été appris sur la partie concepts de ce corpus.

Dans un premier temps, le graphe FST/CRF est obtenu en composant tous les modèles $\lambda_S \circ \lambda_R \circ \lambda_T \circ \lambda_F \circ \lambda_L$ comme cela a été proposé pour la traduction. Cette approche donne un CER de 15,3% qui est bien moins bon que l'approche CRFs de base (12,9%).

Afin de pouvoir se rapprocher du modèle CRFs (qui ne comprend pas un modèle de réordonnancement ni un modèle de langage), nous proposons d'obtenir le graphe FST/CRF en combinant uniquement les modèles $\lambda_S \circ \lambda_T \circ \lambda_F$. Cela nous a permis d'augmenter la performance de cette approche de 1,6% absolu (15,3% vs 13,7%).

Une comparaison entre la performance de ce modèle et la performance du modèle CRFs de base est donnée dans le tableau 7.9. Les résultats montrent qu'en utilisant l'approche FST/CRF (sans modèles de langage et de réordonnancement) nous arrivons à retrouver presque les mêmes performances (12,9% vs. 13,7%) que celles obtenues par l'approche CRFs connue pour être la plus performante pour une tâche de compréhension.

Enfin, vu que l'approche FST/CRF a tendance à projeter les mots hors-vocabulaire en sortie et que cela (bien qu'il soit avantageux pour la traduction) est pénalisant dans une tâche de compréhension, nous proposons de faire un traitement spécifique des mots hors-vocabulaire de la même manière que cela a été proposé auparavant. Ce traitement permet d'augmenter les performances obtenues de 0,3% absolu (13,7% vs. 13,4%).

7.5 Décodage conjoint dans le cas d'un scénario de portabilité du français vers l'italien d'un système de compréhension (TestOnSource)

Un décodage conjoint pour la traduction et la compréhension a été appliqué comme nous l'avons proposé dans la section 6.4. Ce décodage consiste à donner le graphe de traduction en entrée du FST/CRF de compréhension (incluant les scores pondérés re-

Couplage	SMT	SLU	BLEU(%)	CER(%)	Oracle
graphe/graphe	FST/CRF	FST/CRF	43,84	21,4	20,6
1-best/graphe	FST/CRF	FST/CRF	44,09	21,8	21,1
1-best/1-best	FST/CRF	CRFs	44,09	21,5	21,2
1-best/graphe	PB-SMT	FST/CRF	47,18	20,4	19,9
1-best/1-best	PB-SMT	CRFs	47,18	19,9	19,8

TABLE 7.10 – *Evaluation des différents modèles conjoints, variant selon le type d'information transmise entre les 2 étapes (1-best ou graphe).*

latifs à la traduction) et ensuite récupérer en sortie un graphe de compréhension qui intègre les scores de traduction et de compréhension.

Pour cela nous avons adapté l'accepteur du modèle de compréhension du français (décrit dans 7.4) pour prendre des graphes en entrée (au lieu d'une chaîne). Ce modèle prend en entrée le graphe de sortie du modèle FST/CRF de traduction italien vers français (décrit aussi dans la section 7.4). Ce transducteur génère un graphe valué de compréhension en gardant en mémoire les scores de traduction.

Au moment du décodage les deux scores (traduction et compréhension) sont pris en considération. Dans un premier temps nous proposons que le score final pour chaque chemin du graphe soit l'addition simple du score de traduction et du score de compréhension sur ce chemin. Le meilleur chemin est ensuite sélectionné parmi l'ensemble des chemins possibles dans le graphe. Ce chemin représente donc un décodage conjoint entre la traduction et la compréhension (marginalisation de la variabilité liée à la traduction intermédiaire).

Afin de pouvoir se positionner par rapport aux autres approches, nous proposons de comparer la performance de ce modèle avec trois autres techniques : dans la première, le meilleur chemin (1-best) du modèle FST/CRF de traduction (et non pas un graphe de traduction) est donné en entrée du système FST/CRF de traduction. Dans la deuxième, la meilleur chemin du modèle FST/CRF de traduction est donnée en entrée d'un modèle CRFs de base et dans la troisième, la meilleure hypothèse du modèle PB-SMT de traduction est donnée en entrée du FST/CRF de compréhension. Le résultat de cette comparaison est donné dans la tableau 7.10.

Ces résultats montrent que le graphe de traduction améliore la performance du système par rapport au système de 1-meilleure traduction (CER 21,4% vs. 21,8%). On remarque aussi que le décodage conjoint en utilisant un graphe de traduction est moins performant que celui qui utilise la meilleure traduction du modèle PB-SMT. Cela peut s'expliquer par la différence entre la performance du système de traduction utilisé dans le modèle conjoint (44,09% de score BLEU) et celle du modèle PB-SMT (47,18% de score BLEU).

La performance du modèle conjoint (graphe/graphe) est assez comparable à celle obtenue par une cascade des 1-meilleurs hypothèse des systèmes PB-SMT et CRFs, nommée auparavant TestOnSource (21,4% vs. 19,9%). Malgré la dégradation de perfor-

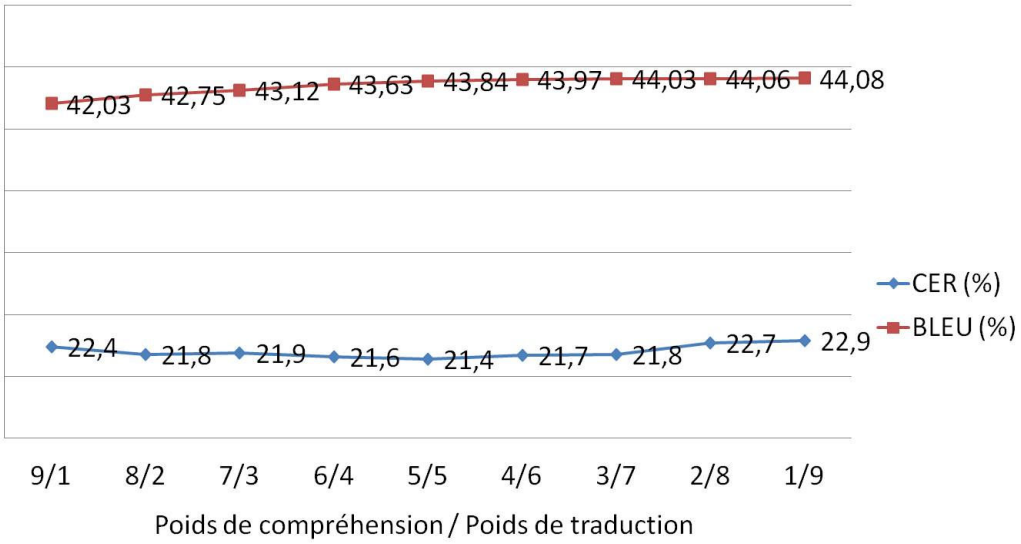

FIGURE 7.3 – *Performances des modules de traduction (% BLEU) et de compréhension (% CER) obtenues par un décodage conjoint en fonction du poids associé à chacun des scores.*

mances, le modèle conjoint garde l'avantage d'être homogène et nous permet d'obtenir un système combiné.

Dans le décodage conjoint qu'on vient de présenter, les scores de traduction et de compréhension ont des poids équivalents (le score global est une addition de ces deux scores). Il est possible que cette jointure simple ne soit pas la meilleure et qu'en donnant des poids plus importants à un des modèles nous puissions améliorer la performance du système.

Pour l'instant, nous n'avons pas un moyen direct pour trouver le poids optimal pour chaque module. Pour cela nous avons proposé d'évaluer les performances (CER et BLEU) que nous pouvons obtenir en donnant des poids différents (fixés a priori) aux modèles. Ainsi le décodage conjoint pourra donc être présenté comme :

$$\lambda_{conjoint} = \Phi_{tradcution}\lambda_{tradcution} \circ \Phi_{comprehension}\lambda_{comprehension}$$

où $\Phi_{traduction}$ est le poids associé au modèle $\lambda_{tradcution}$ et $\Phi_{comprehension}$ est le poids associé au modèle $\lambda_{comprehension}$

Les résultats de cette évaluation sont présentés dans la figure 7.3. Les résultats sur les différents jeux de poids montrent que le score BLEU augmente toujours lorsqu'on augmente le poids du modèle de traduction, par contre l'augmentation des poids du modèle de compréhension n'améliore par forcement le CER.

Ces résultats préliminaires montrent clairement que l'ajustement des poids pour un tel décodage conjoint est un problème délicat auquel il faudra apporter une solution plus optimale.

D'autre part, la performance de ce modèle conjoint peut être améliorée en améliorant le système de traduction, notamment en utilisant un modèle de réordonnancement plus consistant par exemple. Les améliorations de ce système feront partie de nos perspectives (voir section 8.2).

Il est important de mentionner que le modèle conjoint a l'avantage d'être basé sur des FST, et donc toutes les suggestions d'améliorations ou de combinaison peuvent être intégrées facilement en utilisant les bibliothèques de l'outil OpenFst. Ceci pausera probablement des problèmes de calculabilité (lorsqu'il s'agit de très grands graphes) et donc des élagages devront être envisagés dans ce cas.

Dans un contexte de dialogue un décodage conjoint entre la reconnaissance de la parole et la traduction peut être appliqué. Dans ce cas un graphe de reconnaissance pourra être composé avec un graphe de compréhension. Cette composition permettra au système de reconnaissance d'envoyer des informations plus riches au système de compréhension comme dans (Minescu et al., 2007) et le système de compréhension transmettra à son tour des informations riches au gestionnaire de dialogue ce qui influencera positivement sur la performance du système global.

7.6 Conclusion

Dans ce chapitre nous avons évalué et comparé une approche discriminante avec une approche générative à la fois pour la compréhension de la parole et pour la traduction automatique.

Nous avons montré que l'approche CRFs reste la meilleure approche pour la compréhension de la parole, malgré toutes les adaptations de l'approche PB-SMT pour la tâche. Cependant, nous avons montré que la combinaison des sorties des modèles CRFs avec des sorties d'un modèle PB-SMT peut améliorer significativement la performance obtenue et donc les modèles sont assez complémentaires.

L'utilisation d'une approche CRFs pour la traduction à plusieurs limites et les performances de cette approche peuvent être améliorées par un passage vers un modèle FST/CRF. Ce modèle permet d'obtenir un graphe de traduction en combinant plusieurs modèles (réordonnancement, traduction, langage). La performance obtenue par cette approche est légèrement moins bonne que celle d'une approche PB-SMT mais elle reste avantageuse pour un décodage conjoint. Plusieurs améliorations peuvent être envisagées pour cette approche, notamment sur le modèle de réordonnancement, le passage vers un modèle génératif par exemple.

De la même manière, nous avons évalué la performance d'un modèle FST/CRF pour la compréhension. Ce modèle à l'avantage de pouvoir générer des graphes en sortie et donc permet une transmission d'informations plus riche vers les niveaux supérieurs dans le système de dialogue (composition sémantique, gestionnaire de dialogue...).

Nous avons évalué notre proposition de décodage conjoint entre la traduction et la

compréhension dans le cas de l'approche de portabilité TestOnSource (déjà présentée dans des chapitres précédents) et nous avons montré qu'avec un tel décodage nous pouvons obtenir de bonnes performances (assez comparables à l'approche TestOnSource initiale) tout en proposant un système homogène avec possibilité de composition avec d'autres systèmes (reconnaissance de la parole par exemple).

Chapitre 8

Conclusions et Perspectives

Sommaire

8.1 Conclusion 132
8.2 Perspectives 134

8.1 Conclusion

Les travaux présentés dans cette thèse peuvent être classés dans deux lignes principales. Dans la première nous nous intéressons à la portabilité multilingue d'un système de compréhension de la parole. Dans cette optique nous proposons d'utiliser la traduction automatique afin de minimiser le coût du développement d'un nouveau système de compréhension dans une nouvelle langue. Dans un second temps, nous cherchons à comparer les approches de traduction automatique avec les approches de compréhension de la parole, afin d'obtenir un décodage conjoint entre la traduction et la compréhension et d'améliorer le fonctionnement des modèles appliqués à chacune des tâches.

La portabilité multilingue d'un système de compréhension automatique de la parole

Dans cette thèse, nous avons proposé plusieurs approches pour la portabilité multilingue d'un système de compréhension de la parole. Ces approches diffèrent essentiellement par le niveau auquel la portabilité est appliquée.

La première approche TestOnSource consiste à porter un système de compréhension au niveau du décodage (test). Notre proposition consiste à obtenir une traduction automatique des entrées de la nouvelle langue et ensuite d'étiqueter ces traductions par un étiqueteur existant dans la langue source.

La deuxième approche TrainOnTarget consiste à porter le système au niveau de l'apprentissage et donc apprendre un nouveau système de compréhension dans la langue cible. Pour cela, nous avons proposé de traduire le corpus d'apprentissage existant vers la langue cible et ensuite nous avons proposé plusieurs approches pour porter l'annotation de ce corpus vers le corpus traduit en se basant sur les techniques de la traduction automatique, notamment l'alignement automatique des mots.

Nous avons évalué et comparé nos propositions en utilisant un système de traduction spécialisé (conçu pour la tâche) et un système de traduction générique disponible en ligne. Ces approches ont été évaluées aussi bien sur la portabilité d'un système de compréhension vers une langue voisine (français vers italien) que sur la portabilité vers une langue de structure très différente (français vers arabe).

Nous avons montré que la méthode TestOnSource est à la fois la plus performante et la plus simple. Cette méthode peut être appliquée directement sur toutes les langues à partir du moment qu'un traducteur automatique de la langue source vers cette langue suffisamment performant est disponible. Cependant les performances de cette méthode sont influencées par le bruit de la traduction automatique ce qui nous a motivé à proposer des approches pour rendre cette méthode plus robuste aux erreurs de traduction.

Afin de rendre l'approche TestOnSource plus robuste aux erreurs de traduction nous avons proposé de post-éditer les sorties de traduction avant de les passer en entrée

du système de compréhension et aussi d'intégrer du bruit de traduction automatique dans les données d'apprentissage du système de compréhension. Nous avons montré que la mise en série de ces méthodes peut augmenter significativement la performance de cette approche.

Nous avons aussi montré que les approches de portabilité peuvent avoir un rôle important dans une procédure d'annotation d'un nouveau corpus. Nous avons proposé d'utiliser une approche de portabilité lors de l'annotation d'un corpus pour produire une pré-annotation automatique qui sera ensuite corrigée par des annotateurs humains. Cette pré-annotation a permis d'obtenir des gains de productivité d'environ 50%en temps humain.

Approches conjointes pour la traduction et la compréhension

Dans la deuxième partie de la thèse nous cherchons à obtenir un décodage conjoint entre la traduction et la compréhension afin de pouvoir combiner les deux tâches au sein d'un seul modèle.

Pour cela, nous avons proposé d'étudier les différences et les relations entre les approches de traduction et les approches de compréhension. Afin de réaliser cette comparaison, nous avons proposé d'utiliser une approche de traduction (PB-SMT) pour effectuer une tâche de compréhension et réciproquement d'utiliser une approche d'étiquetage (CRFs) utilisée pour la compréhension pour réaliser la traduction automatique.

Ainsi, la compréhension de la parole est considérée comme une traduction d'une chaîne de mots vers une chaîne de concepts et la traduction automatique est considérée comme un étiquetage sémantique des phrases de la langue source par des étiquettes qui correspondent au vocabulaire de la langue cible. Malgré cette représentation, chaque tâche possède des spécificités qui doivent être prises en compte afin de pouvoir optimiser ses performances.

Nous avons montré que ces deux approches diffèrent dans la manière de gérer le réordonnancement, dans l'alignement entre les entrées et les sorties, dans l'utilisation du modèle de langage ou encore dans le traitement des entrèes inconnues (mots hors vocabulaire). Les caractéristiques de chaque tâche doivent être prises en compte pour obtenir de bonnes performances.

Malgré toutes les adaptations possibles sur l'approche PB-SMT, ses performances restent moins bonnes que celles de l'approche CRFs pour la compréhension. D'autre part l'utilisation des CRFs pour la traduction présente plusieurs limites notamment dans la prise en compte du réordonnancement et du modèle de langage.

Pour obtenir un décodage conjoint entre un modèle de compréhension et un modèle de traduction, ces modèles doivent être représentés d'une manière homogène. Pour cela nous proposons de recourir à une approche à base de graphes à états finis (FST) qui permet d'obtenir un graphe pour la traduction et un autre pour la compréhension. La composition de ces deux graphes représente donc un décodage conjoint par marginalisation de l'étape de traduction intermédiaire. Un tel décodage ne cherche plus à obtenir

la meilleure traduction mais la traduction qui permet d'aboutir à la meilleure interprétation possible de l'entrée.

Nous avons montré que cette approche permet d'obtenir des performances très comparables à celles obtenues par la méthode TestOnSource en ayant l'avantage d'être sous forme de graphe. La représentation en graphe permet, dans le cadre d'un système de dialogue, de récupérer des informations riches du module de compréhension (sous forme de graphe) et de transférer des informations riches au gestionnaire de dialogue.

8.2 Perspectives

Les travaux effectués dans cette thèse ouvrent plusieurs perspectives, que ce soit au niveau du multilinguisme du système de compréhension ou au niveau du décodage conjoint entre plusieurs approches.

Perspectives pour la portabilité

L'évaluation des approches de portabilité présentées dans cette thèse était basé sur un corpus de test transcrit manuellement. Cette évaluation permet d'avoir une comparaison entre les différentes approches mais ne permet pas d'avoir une estimation de la performance du système de compréhension dans le cadre d'un système de dialogue.

Une perspective de ce travail serait de se mettre dans des conditions réelles d'un système de dialogue et d'évaluer nos propositions de portabilité sur les sorties d'un système de reconnaissance au lieu des entrées transcrites.

De la même manière que nous l'avons proposé pour augmenter la robustesse du système de compréhension aux erreurs de traduction il serait intéressant d'appliquer ces méthodes pour mieux gérer le bruit venant de la transcription. Cela peut être réalisé en post-éditant les sorties de reconnaissance et en rajoutant des données bruitées dans le corpus d'apprentissage.

Dans toutes les approches proposées pour la portabilité, nous avons utilisé des systèmes de traduction automatique. Les meilleures performances ont été obtenues en utilisant des systèmes appris sur des données du domaine.

Toutefois, nos efforts cherchaient à trouver la meilleure méthode pour porter un système de compréhension vers une nouvelle langue sans chercher à améliorer ou adapter le système de traduction.

Une perspective intéressante à ce travail serait de proposer des solutions pour adapter le système de traduction aux besoins de la portabilité. Une des voies possibles est d'utiliser des systèmes de traduction plus complexes (hiérarchiques, tree-based ou factorisés, par exemple) et d'essayer de trouver des liens entre la représentation hiérarchique de la phrase et sa représentation sémantique afin d'aider la traduction. Il sera alors aussi possible d'enrichir la traduction par des informations sémantiques.

Comme nous l'avons montré dans cette thèse, les méthodes fondées sur la portabilité du corpus d'apprentissage vers une langue comme l'arabe ne sont pas applicables directement.

Des travaux, comme (Misu et al., 2012), ont proposé des méthodes adaptées pour à la portabilité d'un système de compréhension vers une langue éloignée. Dans la même optique une perspective de cette thèse est de proposer des solutions efficaces pour pouvoir appliquer la portabilité vers des langues qui ont un ordre de mots assez différent ou qui ont une riche représentation morphologique.

Pour cela des solutions peuvent être apportées en utilisant par exemple des prétraitements ou des post-traitements statistiques pour effectuer le réordonnancement et des techniques de traduction automatique pour guider la segmentation.

Une autre perspective consiste à étudier l'influence des différentes traductions d'un énoncé sur sa compréhension. Pour cela notre proposition consiste à traduire l'énoncé à étiqueter vers n langues différentes et ensuite étiqueter les n versions par n étiqueteurs sémantiques (obtenus par des approches de portabilité par exemple). Les différents étiquetages de cette même phrase peuvent être regroupés dans un réseau de confusion et la meilleure hypothèse peut être sélectionnée.

L'idée derrière cette proposition est de vérifier si certaines phrases peuvent être mieux étiquetées dans une langue que dans une autre et aussi si la combinaison de l'étiquetage des différentes traductions d'une même phrase peut améliorer le système.

Perspective pour le décodage conjoint

Dans le chapitre 7 nous avons montré que l'utilisation d'une approche conjointe pour la traduction et la compréhension donne des bonnes performances. Cette approche est basée sur la composition d'un graphe de traduction avec un graphe de compréhension. En terme de perspective nous proposons de composer ce graphe avec un graphe de sortie de reconnaissance automatique. Cela permettra d'optimiser le choix de la meilleure transcription en fonction de l'étiquetage qui suivra.

Un problème de taille de graphe se posera dans une telle combinaison et donc une méthode d'élagage doit être proposée afin de diminuer la taille du graphe en éliminant les chemins de plus faible scores. Cela peut être fait d'une manière similaire à l'élagage utilisé lors du décodage d'un système de traduction à base de segment (voir section 3.3.4).

Nous avons montré également dans le chapitre 7 que la performance du modèle conjoint varie selon les poids qu'on attribue à chaque composant de ce modèle. L'ajustement de ces poids permettra d'optimiser la performance globale du système.

Cette optimisation peut être appliquée à deux niveaux : le premier consiste à optimiser les poids internes dans chaque modèle (réordonnancement, modèle de langage, modèle de traduction) et le deuxième consiste à donner des poids aux différents composants du décodage conjoint.

Le décodage conjoint pourra donc être présenté comme :

$$\lambda_{conjoint} = \Phi_{tradcution}\lambda_{tradcution} \circ \Phi_{comprehension}\lambda_{comprehension}$$

où Φ_i est le poids associé au modèle i,

$$\lambda_{traduction} = \Phi_S\lambda_S \circ \Phi_R\lambda_R \circ \Phi_T\lambda_T \circ \Phi_F\lambda_F \circ \Phi_L\lambda_L$$

et

$$\lambda_{comprehension} = \Phi_S\lambda_S \circ \Phi_R\lambda_R \circ \Phi_T\lambda_T \circ \Phi_F\lambda_F \circ \Phi_L\lambda_L$$

Les différents poids doivent être ajustés afin d'obtenir un décodage optimal. Cette optimisation peut être réalisée en utilisant un algorithme similaire à MERT (voir le tableau 3.1).

Le modèle de traduction FST/CRF utilisé dans le décodage conjoint est moins performant que le modèle PB-SMT. Plusieurs améliorations peuvent être appliquées sur ce modèle afin de pouvoir augmenter sa performance.

Une voie d'amélioration possible est d'avoir un modèle réordonnancement plus performant que celui utilisé jusqu'à présent. Le modèle utilisé dans les expériences présentées dans cette thèse est basé sur un ensemble de règles extraites à partir d'un corpus d'apprentissage.

Nous pensons que le passage vers un modèle discriminant pour le réordonnancement peut améliorer la performance du système. Cette proposition consiste à apprendre un modèle CRFs pour le réordonnancement et ensuite le représenter sous forme de graphe. De cette manière nous obtenons un graphe de réordonnancement qui sera composé avec les autres composants du modèle de traduction afin de réaliser un réordonnancement optimal.

Liste des illustrations

1.1 Architecture générale d'un système de dialogue. 11
1.2 Illustration du projet ANR PORT-MEDIA : le processus de production de nouvelles données. 15

2.1 Architecture générale d'un système de dialogue. 20
2.2 Arbre sémantique associé à la phrase "je voudrais réserver un hôtel à Paris le 5 Juin". 27
2.3 Représentation graphique des modèles HMM, MEMM et CRFs. 30
2.4 Projection des données dans un espace de grande dimension. 33

3.1 Le triangle de Vauquois. 39
3.2 Exemple d'un alignement de mots. 45
3.3 Exemple d'alignement produit par les modèle IBM-1, IBM-2 et HMM . . 47
3.4 Exemple d'alignement produit par les modèle IBM-3, IBM-4 et IBM-5. . 47
3.5 Un alignement symétrisé obtenu par l'union d'alignements en mots. . . 50
3.6 Exemple d'extraction de segments 51
3.7 L'algorithme de décodage et les piles des hypothèses pour la recherche en faisceaux. 54

4.1 La méthode TestOnSource. 68
4.2 La méthode TrainOnTarget. 69
4.3 Exemple de projection d'étiquettes sémantiques en utilisant un alignement direct (non-supervisé). 70
4.4 Exemple d'erreur en utilisant un alignement (non-supervisé) pour la projection de concepts. 70
4.5 Exemple de projection de concepts sémantiques en utilisant un alignement indirect (semi-supervisé). 72
4.6 Exemple d'une situation ambiguë pour la projection de concepts. 72
4.7 Accroître la robustesse de la méthode TestOnSource en utilisant des données d'apprentissage bruité. 74
4.8 La mise en série des méthodes proposées pour accroître la robustesse de la méthode TestOnSource aux erreurs de traduction. 75

5.1 Protocole du Magicien d'Oz. 79

5.2 L'hypothèse générée par le modèle CRFs pour la phrase "je voudrais réserver un hôtel à Nice" (le mot "Nice" est hors vocabulaire). 84
5.3 L'hypothèse générée par le modèle CRFs pour la phrase "Y-a-t il une autoroute près de l'hôtel" (le mot "autoroute" est hors vocabulaire). 84
5.4 L'hypothèse générée par le modèle CRFs pour la phrase "una camera dopia con un letto per bambino" traduite par "une chambre double avec un lit" 86
5.5 L'hypothèse générée par le modèle CRFs pour la phrase "una camera dopia con un letto per bambino" traduite par "une chambre double avec un lit" ensuite post-éditée par "une chambre double avec un lit enfant". 88
5.6 Exemple d'inadéquation en utilisant la méthode d'alignement. 92
5.7 Exemple d'un alignement ambigu qui génère des erreurs sémantiques. . 92
5.8 Pourcentage de gain de productivité de l'annotation sémantique. 94

6.1 Exemple d'extraction de règle de réordonnancement 108
6.2 Exemple de graphe composé de plusieurs chemins de réordonnancement 109
6.3 Exemple de feature matcher. Le symbole (*) représente toute observation possible. 110

7.1 Comparaison entre l'hypothèse générée par le modèles SLU/CRF et celle générée par le modèle SLU/PB-SMT pour la phrase "Je voudrais réserver en fait pour la ville de Nice du premier aux trois novembre". 120
7.2 Comparaison entre l'hypothèse générée par le modèles SLU/CRF et celle générée par le modèle SLU/PB-SMT pour la phrase "un autre quartier du côté de Saint-Michel". 120
7.3 Performances des modules de traduction (% BLEU) et de compréhension (% CER) obtenues par un décodage conjoint en fonction du poids associé à chacun des scores. 128

Liste des tableaux

2.1 Une comparaison entre des corpus de dialogue pour des domaines et des langues différents. . . . 25
2.2 Exemple de cadre sémantique pour la tâche ATIS. . . . 28

3.1 Le pseudo code de l'algorithme MERT pour l'optimisation du modèle log-linéaire. . . . 52
3.2 Le pseudo code de l'algorithme de recherche en faisceau. . . . 53

5.1 Un extrait de dialogue du corpus MEDIA. . . . 80
5.2 Caractéristiques du corpus MEDIA 80
5.3 Exemple d'annotation sémantique du corpus MEDIA. . . . 81
5.4 Evaluation des systèmes de traduction. . . . 82
5.5 Aperçu du corpus MEDIA et de sa traduction vers l'italien (# phrases). . . . 82
5.6 Evaluation (CER %) du SLU français de référence. . . . 83
5.7 Evaluation (CER %) des différentes stratégies de portabilité d'un système de compréhension de l'italien. . . . 85
5.8 Evaluation (CER %) des méthodes de portabilité appliquées sur 5.6k phrases traduites manuellement. . . . 86
5.9 Evaluation (CER %) des approches proposées pour la robustesse des systèmes au bruit de traduction. . . . 87
5.10 combinaison de systèmes 89
5.11 Evaluation (CER %) des différentes stratégies de portabilité en utilisant des traductions en ligne. . . . 90
5.12 Evaluation (CER %) de la méthode TestOnSource pour l'italien et l'arabe. 93
5.13 Le gain de productivité% et le CER% pour chaque itération du processus d'annotation sémantique. L'ensemble de test PMEDIA utilisé pour mesurer les CER a été compilé après le processus d'annotation avec des données de chaque bloc. . . . 94
5.14 Les CER(%) obtenues par le modèle MEDIA sur les différents blocs d'annotation du corpus PMEDIA. . . . 95
5.15 Evaluation (CER %) des modèles italiens sur deux ensembles de test différents. . . . 96

6.1 Comparaison entre les approches utilisées en traduction (SMT) et en compréhension de la parole (SLU). . . . 114

7.1 Les améliorations itératives du modèle SLU/PB-SMT sur le corpus MEDIA français(CER%). 119
7.2 Comparaison des types d'erreurs entre l'approche CRFs et l'approche PB-SMT (CER%). 119
7.3 Les améliorations itératives du modèle SLU/PB-SMT sur le corpus MEDIA italien(CER%). 121
7.4 Combinaison des système de compréhension de l'italien, avec et sans l'approche PB-SMT. 121
7.5 Evaluation du modèle CRFs pour la traduction du français vers l'italien (BLEU %). 122
7.6 Comparaison objective entre le modèle PB-SMT et le modèle CRFs pour la traduction du français vers l'italien (BLEU %). 123
7.7 Comparaison entre le modèle PB-SMT et le modèle CRFs pour la traduction de l'italien vers le français (BLEU %). 124
7.8 Comparaison entre les différentes approches (PB-SMT, CRFs, FST/CRF) pour la traduction. 125
7.9 Evaluation de l'approche FST/CRF pour la compréhension du français. 126
7.10 Evaluation des différents modèles conjoints, variant selon le type d'information transmise entre les 2 étapes (1-best ou graphe). 127

8.1 Evaluation (CER %) des modèles français sur le test PM-DOM. 158

Bibliographie

Akbacak, M., Gao, Y., Gu, L., and Kuo, H. (2005). Rapid transition to new spoken dialogue domains : Language model training using knowledge from previous domain applications and web text resources. In *The European Conference on Speech Communication and Technology (EUROSPEECH)*, pages 1873–1876, Lisbon, Portugal. ISCA.

Al-onaizan, Y., Curin, J., Jahr, M., Knight, K., Lafferty, J., Melamed, D., Och, F.-J., Purdy, D., Smith, N. A., and Yarowsky, D. (1999). Statistical machine translation. Technical report, Final Report, JHU Summer Workshop.

Allauzen, A. and Wisniewski, G. (2010). Modèles discriminants pour l'alignement mot-à-mot. *Traitement Automatique des Langues (TAL)*, 50(3/2009) :173–203.

Allauzen, C., Riley, M., Schalkwyk, J., Skut, W., and Mohri, M. (2007). OpenFst : A general and efficient weighted finite-state transducer library. In *The International Conference on Implementation and Application of Automata, (CIAA)*, volume 4783 of *Lecture Notes in Computer Science*, pages 11–23. Springer. `http://www.openfst.org`.

Allen, J. F. (1987). *Natural language understanding*. Bejnamin/Cummings series in computer science. Addison-Wesley.

Allen, J. F., Schubert, L. K., Ferguson, G., Heeman, P., Hwang, C. H., Kato, T., Light, M., Martin, N. G., Miller, B. W., Poesio, M., and Traum, D. R. (1994). The TRAINS project : A case study in defining a conversational planning agent. Technical Report TN94-3, University of Rochester, Computer Science Department. Thu, 17 Jul 97 09 :00 :00 GMT.

Anoop Deoras, R. S. G. T. and Hakkani-Tur, D. (2012). Joint decoding for speech recognition and semantic tagging. In *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Portland, USA. ISCA.

Arnold, D., Balkan, L., Humphries, R. L., Meijer, S., and Sadler, L. (1993). Machine translation : an introductory guide. *NCC Blackwell*, 4(4) :363–382.

Aust, H., Oerder, M., Seide, F., and Steinbiss, V. (1995). The Philips automatic train timetable information system. *Speech Communication*, 17 :249–262.

Bahl, L. R., Baker, J. K., Jelinek, F., and Mercer, R. L. (1977). Perplexityâa measure of the difficulty of speech recognition tasks. *Journal of The Acoustical Society of America*, 62 :63.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The berkeley framenet project. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, pages 86–90, Montreal, Canada. ACL.

Béchara, H., Ma, Y., and van Gebabith, J. (2011). Statistical post-editing for a statistical machine translation system. In *Machine Translation Summit Conference*, pages 308–317, Xiamen, China. AMTA.

Bennacef, S., Bonneau-Maynard, H., Gauvain, J.-L., Lamel, L., and Minker, W. (1994). A spoken language system for information retrieval. In *The International Conference on Spoken Language Processing (ICSLP)*, Yokohama, Japan. ISCA.

Bennacef, S., Devillers, L., Rosset, S., and Lamel, L. (1996). Dialog in the RAILTEL telephone-based system. In *The International Conference on Spoken Language Processing (ICSLP)*, volume 1, pages 550–553, Philadelphia, PA. ISCA.

Bohus, D., Raux, A., Harris, T., Eskenazi, M., and Rudnicky, A. (2007). Olympus : an open-source framework for conversational spoken language interface research. In *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLTNAACL)*, pages 32–39, Rochester, New York. ACL.

Boitet, C. (2008). Les architectures linguistiques et computationnelles en traduction automatique sont indépendantes. In *TALN*, Avignon, France. ATALA.

Bonneau-Maynard, H. and Lefèvre, F. (2001). Investigating stochastic speech understanding. In *The Workshop on Automatic Speech Recognition and Understanding (ASRU)*, Trento, Italy. IEEE.

Bonneau-Maynard, H. and Rosset, S. (2003). A semantic representation for spoken dialogs. In *The European Conference on Speech Communication and Technology (EUROSPEECH)*, pages 253–256, Geneva, Switzerland. ISCA.

Bonneau-Maynard, H., Rosset, S., Ayache, C., Kuhn, A., and Mostefa, D. (2005). Semantic annotation of the french media dialog corpus. In *The European Conference on Speech Communication and Technology (EUROSPEECH)*, pages 3457–3460, Lisbon, Portugal. The International Speech Communication Association.

Bos, J., Klein, E., Lemon, O., and Oka, T. (2003). Dipper : Description and formalisation of an information-state update dialogue system architecture. In *SIGdial Workshop on Discourse and Dialogue*, pages 115–124, Sapporo, Japan. ACL.

Brachman, R. J. (1979). On the epistemological status of semantic networks. In Findler, N. V., editor, *Associative Networks : Representation of Knowledge and Use of Knowledge by Examples*. Academic Press, New York.

Brown, P. F., Cocke, J., Pietra, S. D., Pietra, V. J. D., Jelinek, F., Lafferty, J. D., Mercer, R. L., and Roossin, P. S. (1990). A statistical approach to machine translation. *Computational Linguistics*, 16(2) :79–85.

Brown, P. F., Pietra, S. D., Pietra, V. J. D., and Mercer, R. L. (1993). The mathematics of statistical machine translation : Parameter estimation. *Computational Linguistics*, 19(2) :263–311.

Bruce, B. (1975). Case Systems for Natural Language. *Artificial Intelligence*, 6(4) :327–360.

Burke, C., Doran, C., Gertner, A., Gregorowicz, A., Harper, L., Korb, J., and Loehr, D. (2003). Dialogue complexity with portability ? research directions for the information state approach. In *the HLTNAACL workshop on Research directions in dialogue*, volume 7, pages 13–15, Stroudsburg, USA. ACL.

Chen, S. F. and Goodman, J. (1996). An empirical study of smoothing techniques for language modeling. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, pages 310–318, Santa Cruz, California. ACL.

Cherkassky, V. (1997). The nature of statistical learning theory. *Transactions on Neural Networks*, 8(6) :1564–1564.

Chiang, D. (2007). Hierarchical phrase-based translation. *Computational Linguistics*, 33(2) :201–228.

Chomsky, N. (1957). *Syntactic structures*. The Hague : Mouton.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, 2 :137–167.

Chomsky, N. and Schützenberger, M. P. (1963). The algebraic theory of context-free languages. In *Computer programming and formal systems*, pages 118–161, Amsterdam. North-Holland.

Crego, J. M. and Mariño, J. B. (2006). Improving statistical mt by coupling reordering and decoding. *Machine Translation*, 20(3) :199–215.

Crego, J. M., Yvon, F., and Mariño, J. B. (2011). Ncode : an open source bilingual n-gram smt toolkit. *The Prague Bulletin of Mathematical Linguistics (PBML)*, 96 :49–58.

de Ilarraza, A., Labaka, G., and Sarasola, K. (2008). Statistical postediting : A valuable method in domain adaptation of rbmt systems for less-resourced languages. *Mixing Approaches to Machine Translation (MATMT)*, pages 35–40.

De Mori, R. (1997). *Spoken Dialogues with Computers*. Academic Press, Inc., Orlando, FL, USA.

Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *The ACL Workshop on Human Language Technology (HLT)*, pages 138–145, San Diego, USA. Morgan Kaufmann.

Dorr, B. J., Jordan, P. W., and Benoit, J. W. (1999). A survey of current paradigms in Machine Translation. Technical report, University of Maryland.

Dymetman, M. and Cancedda, N. (2010). The intersecting hierarchical and phrase-based models of translation : Formal aspects and algorithms. In *The Workshop on Syntax and Structure in Statistical Translation*, pages 1–9, Beijing, China. Coling 2010 Organizing Committee.

Federico, M., Stüker, S., Bentivogli, L., Paul, M., Cettolo, M., Herrmann, T., Niehues, J., and Moretti, G. (2012). The iwslt 2011 evaluation campaign on automatic talk translation. In Chair), N. C. C., Choukri, K., Declerck, T., Doğan, M. U., Maegaard, B., Mariani, J., Odijk, J., and Piperidis, S., editors, *The International Conference on Language Resources and Evaluation (LREC)*, Istanbul, Turkey. ELRA.

Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni di semantica*, 4(2) :222–255.

Fohr, D., Mella, O., Cerisara, C., and Illina, I. (2004). The automatic news transcription system : ANTS, some real time experiments. In *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Jeju Island, Korea. ISCA.

Fung, P. and Schultz, T. (2008). Multilingual spoken language processing - challenges for multilingual systems. *Signal Processing Magazine*, (89) :89–97.

Gao, Y., Gu, L., and Kuo, H. (2005). Portability challenges in developing interactive dialogue systems. In *The International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 5, pages 1017–1020, Philadelphia, USA. IEEE.

Gascó i Mora, G. and Sánchez Peiró, J. (2007). Part-of-speech tagging based on machine translation techniques. *Pattern Recognition and Image Analysis*, pages 257–264.

Gauvain, J. L., Lamel, L. F., Adda, G., and Adda-Decker, M. (1994). The limsi continuous speech dictation system. In *The ACL Workshop on Human Language Technology (HLT)*, pages 319–324, Plainsboro, USA. ACL.

Germann, U., Jahr, M., Knight, K., Marcu, D., and Yamada, K. (2001). Fast decoding and optimal decoding for machine translation. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, pages 228–235, Toulouse, France. ACL.

Glass, J. (1999). Challenges for spoken dialogue systems. In *The Workshop on Automatic Speech Recognition and Understanding (ASRU)*, Keystone USA. IEEE.

Glass, J., Flammia, G., Goodine, D., Phillips, M., Polifroni, J., Sakai, S., Seneff, S., and Zue, V. (1995). Multilingual spoken-language understanding in the mit voyager system. *Speech Communication*, 17(1) :1–18.

Gorin, A. L., Riccardi, G., and Wright, J. H. (1997). How may I help you ? *Speech Communication*, 23(1/2) :113–127.

H., S. (1994). Probabilistic part-of-speech tagging using decision trees. In *The International Conference on New Methods in Language Processing (NMLP)*, pages 44–49, Manchester.

Haffner, P., Tur, G., and Wright, J. H. (2003). Optimizing svms for complex call classification. In *The International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 632–635, Hong Kong, China. IEEE.

Hahn, S., Dinarelli, M., Raymond, C., Lefèvre, F., Lehnen, P., De Mori, R., Moschitti, A., Ney, H., and Riccardi, G. (2010). Comparing stochastic approaches to spoken language understanding in multiple languages. *Transactions on Audio, Speech, and Language Processing (TASLP)*, 19(6) :1569–1583.

Hahn, S., Lehnen, P., Heigold, G., and Ney, H. (2009). Optimizing crfs for slu tasks in various languages using modified training criteria. In *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 2727–2730, Brighton, UK. ISCA.

Hahn, S., Lehnen, P., Raymond, C., and Ney, H. (2008). A comparison of various methods for concept tagging for spoken language understanding. In *The International Conference on Language Resources and Evaluation (LREC)*, Marrakech, Morocco. ELRA.

Hakkani-Tür, D. Z., Béchet, F., Riccardi, G., and Tür, G. (2006). Beyond asr 1-best : Using word confusion networks in spoken language understanding. *Computer Speech and Language*, pages 495–514.

Hayashi, K., Tsukada, H., Sudoh, K., Duh, K., and Yamamoto, S. (2010). Hierarchical phrase-based machine translation with word-based reordering model. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, COLING '10, pages 439–446, Beijing, China. ACL.

Hemphill, C. T., Godfrey, J. J., and Doddington, G. R. (1990). The atis spoken language systems pilot corpus. In *The Workshop on Speech and Natural Language*, HLT '90, pages 96–101, Hidden Valley, Pennsylvania. ACL.

Herv, D. C., Chen, D., and Bourlard, H. (2001). Text identification in complex background using svm. In *The International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 621–626, Hawaii, USA. IEEE.

Ho, P. and Moreno, P. J. (2004). Svm kernel adaptation in speaker classification and verification. In *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Jeju Island, Korea. ISCA.

Huang, X., Acero, A., Alleva, F., Hwang, M., Jiang, L., and Mahajan, M. (1995). Microsoft windows highly intelligent speech recognizer : Whisper. In *The International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 93–96, Detroit, USA. IEEE.

Huet, S. and Lefèvre, F. (2011). Unsupervised alignment for segmental-based language understanding. In *The Workshop on Unsupervised Learning in NLP*, EMNLP, pages 97–104, Stroudsburg, PA, USA. ACL.

Hutchins, W. J. (2007). Machine translation : A concise history. *Journal of Translation Studies*, pages 29–70.

Hutchins, W. J. and Somers, H. L. (1992). *An Introduction to Machine Translation*. Academic Press.

Jabaian, B., Besacier, L., and Lefèvre, F. (2010). Investigating multiple approaches for slu portability to a new language. In *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 2502–2505, Makuhari, Japan. ISCA.

Joachims, T. (1998). Text categorization with suport vector machines : Learning with many relevant features. In *European Conference on Machine Learning (ECML)*, pages 137–142, London, UK. Springer.

Kaplan, R. M. and Bresnan, J. (1995). Lexical-functional grammar : A formal system for grammatical representation.

Katz, S. M. (1987). Estimation of probabilities from sparse data for the language model component of a speech recognizer. *Transaction on Acoustics, Speech and Signal Processing (TASSP)*, ASSP-35(3) :400.

Khalilov, M. and Fonollosa, J. A. R. (2009). N-gram-based statistical machine translation versus syntax augmented machine translation : Comparison and system combination. In *The Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 424–432, Athens, Greece. ACL.

Kim, W. and Khudanpur, S. (2003). Language model adaptation using cross-lingual information. In *The European Conference on Speech Communication and Technology (EUROSPEECH)*, Geneva, Switzerland. ISCA.

Kingsbury, P. and Palmer, M. (2003). Propbank : The next level of treebank. In *Proceedings of Treebanks and Lexical Theories*.

Knight, K. (1999). Decoding complexity in word-replacement translation models. *Computational Linguistics*, 25(4) :607–615.

Koehn, P. (2004). Pharaoh : A beam search decoder for phrase-based statistical machine translation models. In *Machine Translation Summit Conference*, pages 115–124, Washington, USA. AMTA.

Koehn, P. (2005). Europarl : A parallel corpus for statistical machine translation. In *Machine Translation Summit Conference*, pages 79–86, Phuket, Thailand. AMTA.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., et al. (2007). Moses : Open source toolkit for statistical machine translation. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, pages 177–180, Prague, Czech Republic. ACL.

Koehn, P., Och, F., and Marcu, D. (2003). Statistical phrase-based translation. In *Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLTNAACL)*, volume 1, pages 48–54, Edmonton, Canada. ACL.

Komatani, K., Katsumaru, M., Nakano, M., Funakoshi, K., Ogata, T., and Okuno, H. (2010). Automatic allocation of training data for rapid prototyping of speech understanding based on multiple model combination. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, pages 579–587, Uppsala, Sweden. ACL.

Komatani, K., Tanaka, K., Kashima, H., and Kawahara, T. (2001). Domain-independent spoken dialogue platform using key-phrase spotting based on combined language model. In *The European Conference on Speech Communication and Technology (EUROSPEECH)*, pages 1319–1322, Aalborg, Denmark. ISCA.

Kroch, A. and Joshi, A. (1985). Linguistic relevance of tree adjoining grammars. Technical report.

Kudo, T. (2005). Crf++ : A crf toolkit. *available on http ://crfpp.googlecode.com/svn/trunk/doc/index.html*.

Kudo, T. and Matsumoto, Y. (2001). Chunking with support vector machines. In *The Conference of the North American Chapter of the Association for Computational Linguistics on Language technologies (NAACL)*, pages 1–8, Pittsburgh, Pennsylvania. ACL.

Kudoh, T. and Matsumoto, Y. (2000). Use of support vector learning for chunk identification. In *The Conference on Computational Natural Language Learning (ConLL)*, pages 142–144, Lisbon, Portugal. ACL.

Kumar, S. and Byrne, W. (2003). A weighted finite state transducer implementation of the alignment template model for statistical machine translation. In *The Conference of the North American Chapter of the Association of Computational Linguistics (NAACL)*, pages 63–70, Edmonton, Canada. ACL.

Lafferty, J., McCallum, A., and Pereira, F. (2001). Conditional random fields : Probabilistic models for segmenting and labeling sequence data. In *The International Conference on Machine Learning (ICML)*, pages 282–289, Williamstown, USA. Morgan Kaufmann.

Lamel, L., Lefevre, F., Gauvain, J., and Adda, G. (2001). Portability issues for speech recognition technologies. In *The International Conference on Human Language Technology Research*, pages 1–7, Stroudsburg, USA. ACL.

Lamel, L., Rosset, S., Gauvain, J. L., and Bennacef, S. (1999). The limsi arise system for train travel information. In *The International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 501–504, Washington, USA. IEEE.

Lamel, L., Rosset, S., Gauvain, J. L., Bennacef, S., Garnier-Rizet, M., and Prouts, B. (2000). The limsi arise system. *Speech Communication*, 31(4) :339–353.

Langlais, P., Yvon, F., and Zweigenbaum, P. (2008). Translating medical words by analogy. In *The Workshop on Intelligent Data Analysis in bioMedicine and Pharmacology (IDAMAP) 2008*, Washington, USA.

Lavergne, T., Cappé, O., and Yvon, F. (2010). Practical very large scale CRFs. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, pages 504–513, Uppsala, Sweden. ACL.

Lavergne, T., Crego, J. M., Allauzen, A., and Yvon, F. (2011). From n-gram-based to crf-based translation models. In *The Workshop on Statistical Machine Translation (WSMT)*, WMT '11, pages 542–553, Edinburgh, Scotland. ACL.

Lavie, A. and Denkowski, M. J. (2009). The meteor metric for automatic evaluation of machine translation. *Machine Translation*, 23(2-3) :105–115.

Lee, C. H., Giachin, E., Rabiner, L. R., Pieraccini, R., and Rosenberg, A. E. (1992). Improved acoustic modeling for large vocabulary continuous speech recognition. *Computer Speech and Language*, 6(2) :103–127.

Lefèvre, F. (2007). Dynamic bayesian networks and discriminative classifiers for multi-stage semantic interpretation. In *The International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 4, pages 13–16, Honolulu, USA. IEEE.

Lefèvre, F., Gauvain, J.-L., and Lamel, L. (2005). Genericity and portability for task-independent speech recognition. *Computer, Speech and Language*, 19(3) :345–363.

Lefèvre, F., Mairesse, F., and Young, S. (2010). Cross-lingual spoken language understanding from unaligned data using discriminative classification models and machine translation. In *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 78–81, Makuhari, Japan. ISCA.

Lefèvre, F., Mostefa, D., Besacier, L., Esteve, Y., Quignard, M., Camelin, N., Favre, B., Jabaian, B., and Rojas-Barahona, L. (2012). Robustness and portability of spoken language understanding systems among languages and domains : the PORT-MEDIA project. In *The International Conference on Language Resources and Evaluation (LREC)*, Istanbul, Turkey. ELRA.

Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady.*, 10(8) :707–710.

Liang, P., Taskar, B., and Klein, D. (2006). Alignment by agreement. In *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLTNAACL)*, pages 104–111, New York, USA. ACL.

Macherey, K., Bender, O., and Ney, H. (2009). Application of statistical machine translation approaches to spoken language understanding. In *The International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 17, pages 803–818, Taipei, Taiwan. IEEE.

Macherey, K., Och, F. J., and Ney, H. (2001). Natural language understanding using statistical machine translation. In Dalsgaard, P., Lindberg, B., Benner, H., and Tan, Z.-H., editors, *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 2205–2208, Aalborg, Denmark. ISCA.

Mairesse, F., Gasic, M., Jurcicek, F., Keizer, S., Thomson, B., Yu, K., and Young, S. (2009). Spoken language understanding from unaligned data using discriminative classification models. In *The International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4749–4752, Taipei, Taiwan. IEEE.

Mangu, L., Brill, E., and Stolcke, A. (2000). Finding consensus in speech recognition : word error minimization and other applications of confusion networks. *Computer, Speech and Language*, 14(4) :373–400.

Marcu, D. and Wong, W. (2002). A phrase-based, joint probability model for statistical machine translation. In *The Conference on Empirical Methods in Natural Language Processing (EMNLP)*, EMNLP '02, pages 133–139, Philadelphia, USA. ACL.

Mariòo, J. B., Banchs, R. E., Crego, J. M., de Gispert, A., Lambert, P., Fonollosa, J. A. R., and Costa-jussà, M. R. (2006). N-gram-based machine translation. *Computational Linguistic*, 32(4) :527–549.

Matrouf, K., Neel, F., Gauvain, J.-L., and Mariani, J. (1989). Adaptive syntax representation in an oral task-oriented dialogue for air-traffic controller training. In *The European Conference on Speech Communication and Technology (EUROSPEECH)*, pages 1187–1190, Paris, France. ISCA.

Maynard, H. and Lefèvre, F. (2002). Apprentissage dâun module stochastique de compréhension de la parole. In *Journées dâEtudes sur la Parole (JEP)*, Nancy. AFCP.

McCallum, A., Freitag, D., and Pereira, F. C. N. (2000). Maximum entropy markov models for information extraction and segmentation. pages 591–598.

McTear, M. F. (2004). *Spoken dialogue technology - toward the conversational user interface*. Springer.

Minescu, B., Damnati, G., Béchet, F., and De Mori, R. (2007). Conditional use of word lattices, confusion networks and 1-best string hypotheses in a sequential interpretation strategy. In *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 1617–1620, Antwerp, Belgium. ISCA.

Minker, W. (1998). *Speech Understanding for Spoken Language Systems - Portability Across Domains and Languages*. Number 2569 in Deutsche Hochschulschriften. Hänsel-Hohenhausen, Frankfurt/Main (Germany).

Minker, W. and Bennacef, S. (2004). *Speech and Human-Machine Dialog*. The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, Boston, USA.

Minker, W., Bennacef, S., and Gauvain, J. L. (1996). A stochastic case frame approach for natural language understanding. In *The International Conference on Spoken Language Processing (ICSLP)*, pages 1013–1016, Philadelphia, USA.

Misu, T., Mizukami, E., Kashioka, H., Nakamura, S., and Li, H. (2012). A bootstrapping approach for slu portability to a new language by inducting unanntated user queries. In *The International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan. IEEE.

Mohri, M. (1997). Finite-state transducers in language and speech processing. *Computational Linguistics*, 23(4) :269–311.

Mohri, M. (2009). Weighted automata algorithms. *Handbook of Weighted Automata*, pages 213–254.

Mohri, M., Pereira, F., and Riley, M. (2002). Weighted finite-state transducers in speech recognition. *Computer Speech and Language*, 16(1) :69–88.

Moore, R. C., tau Yih, W., and Bode, A. (2006). Improved discriminative bilingual word alignment. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, ACL-44, pages 513–520, Sydney, Australia. ACL.

Nagao, M. (1984). A framework of a mechanical translation between English and Japanese by analogy principle. In Elithorn, A. and Banerji, R., editors, *Artifical and Human Intelligence*, pages 173–180. North-Holland.

Nocera, P., Linares, G., Massonié, D., and Lefort, L. (2002). Phoneme lattice based a* search algorithm for speech recognition. In *The International Conference on Text, Speech and Dialogue*, pages 301–308, Brno, Czech Republic. Springer.

Och, F. (2003a). Minimum error rate training in statistical machine translation. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, volume 1, pages 160–167, SAPPORO, JAPAN. ACL.

Och, F. and Ney, H. (2000). Improved statistical alignment models. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, pages 440–447, Hong Kong, China. ACL.

Och, F. J. (2003b). Statistical machine translation : From single-word models to alignment templates. Technical Report AIB-2003-06, RWTH Aachen.

Och, F. J. (2005). Statistical machine translation : Foundations and recent advances.

Och, F. J. and Ney, H. (2002). Discriminative training and maximum entropy models for statistical machine translation. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, ACL '02, pages 295–302, Philadelphia, Pennsylvania. ACL.

Och, F. J. and Ney, H. (2003). A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1) :19–51.

Och, F. J. and Ney, H. (2004). The alignment template approach to statistical machine translation. *Computational Linguistics*, 30(4) :417–449.

Och, F. J., Tillmann, C., Ney, H., and Informatik, L. F. (1999). Improved alignment models for statistical machine translation. In *The Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 20–28. ACL.

Och, F. J., Ueffing, N., and Ney, H. (2001). An efficient a* search algorithm for statistical machine translation. In *Data-Driven Machine Translation Workshop*, pages 55–62, Toulouse, France. ACL.

Papineni, K., Roukos, S., Ward, T., and Zhu, W. (2002). Bleu : a method for automatic evaluation of machine translation. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, pages 311–318, Philadelphia, USA. ACL.

Pereira, F. and Wright, R. (1997). *Finite-State Language Processing*. MIT Press, Cambridge, Massachusetts.

Pieraccini, R., Levin, E., and Lee, C.-H. (1991). Stochastic representation of conceptual structure in the atis task. In *The ACL Workshop on Human Language Technology (HLT)*, pages 121–124, Pacific Grove, California. ACL.

Pinault, F. and Lefèvre, F. (2011). Unsupervised clustering of distributions of semantic frame graphs for pomdp-based spoken dialog systems with summary space. In *The Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Barcelona, Spain.

Pollard, C. (1985). Lecture notes on Head-driven Phrase Structure Grammar. Technical report, Center for the Study of Language and Information, Stanford University, USA.

Quillian, M. R. (1968). Semantic memory. In Minsky, M., editor, *Semantic Information Processing*, pages 227–270. MIT Press, Cambridge, MA.

Rama, T., Singh, A., and Kolachina, S. (2009). Modeling letter-to-phoneme conversion as a phrase based statistical machine translation problem with minimum error rate training. In *Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 90–95, Boulder, USA. ACL.

Ramshaw, L. and Marcus, M. (1995). Text chunking using transformation-based learning. In *The Workshop on Very Large Corpora*, pages 82–94, Cambridge, USA. ACL.

Raymond, C., Béchet, F., De Mori, R., and Damnati, G. (2006). On the use of finite state transducers for semantic interpretation. *Speech Communication*, 48(3-4) :288–304.

Raymond, C. and Riccardi, G. (2007). Generative and discriminative algorithms for spoken language understanding. In *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 1605–1608, Antwerp, Belgium. ISCA.

Riedmiller, M. and Braun, H. (1993). A direct adaptive method for faster backpropagation learning : The RPROP algorithm. In *The International Conference on Neural Networks (ICNN)*, pages 586 – 591, San Francisco, USA. IEEE.

Roark, B., Saraclar, M., and Collins, M. (2007). Discriminative n-gram language modeling. *Computer Speech and Language*, 21(2) :373–392.

Roche, E. and Schabes, Y. (1995). Deterministic part-of-speech tagging with finite-state transducers. *Computational Linguistics*, 21(2) :227–253.

Sadek, M., Ferrieux, A., Cozannet, A., Bretier, P., Panaget, F., and Simonin, J. (1996). Effective human-computer cooperative spoken dialogue : The AGS demonstrator. In *The International Conference on Spoken Language Processing (ICSLP)*, volume 1, pages 546–549, Philadelphia, PA. ISCA.

Sarikaya, R. (2008). Rapid bootstrapping of statistical spoken dialogue systems. *Speech Communication*, 50(7) :580–593.

Schapire, R. E. and Singer, Y. (2000). Boostexter : A boosting-based system for text categorization. *Machine Learning*, 39(2/3) :135.

Schultz, T. (2004). Towards rapid language portability of speech processing systems. In *The Conference on Speech and Language Systems for Human Communication*, Delhi, India.

Schultz, T. and Black, A. (2006). Challenges with rapid adaptation of speech translation systems to new language pairs. In *The International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 5, Toulouse, France. IEEE.

Schwartz, R., Miller, S., Stallard, D., and Makhoul, J. (1996). Language understanding using hidden understanding models. In *The International Conference on Spoken Language Processing (ICSLP)*, volume 2, pages 997–1000, Philadelphia, USA. ISCA.

Seneff, S. (1989). TINA : a probabilistic syntactic parser for specch understanding system. In *the DARPA Speech and Natural Language Workshop*.

Seneff, S. and Wang, C. (1997). *Porting the galaxy system to Mandarin Chinese*. Massachusetts Institute of Technology.

Servan, C., Camelin, N., Raymond, C., Béchet, F., and De Mori, R. (2010). On the use of machine translation for spoken language understanding portability. In *The International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5330–5333, Dallas, USA. IEEE.

Servan, C., Raymond, C., Béchet, F., and Nocera, P. (2006). Conceptual decoding from word lattices : application to the spoken dialogue corpus MEDIA. In *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 1614–1617, Pittsburgh, USA. ISCA.

Simard, M. (1998). The baf : A corpus of english-french bitext. In *The International Conference on Language Resources and Evaluation (LREC)*, Granada, Spain. ELRA.

Simard, M., Goutte, C., Isabelle, P., et al. (2007). Statistical phrase-based post-editing. In *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLTNAACL)*, pages 508–515, Rochester, USA. ACL.

Siu, K. and Meng, H. (1999). Semi-automatic acquisition of domain-specific semantic structures. In *The European Conference on Speech Communication and Technology (EUROSPEECH)*, pages 172–181, Budapest, Hungary. ISCA.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Machine Translation Summit Conference*, pages 223–231, Cambridge USA. AMTA.

Song, Fei, Croft, and Bruce, W. (1999). A general language model for information retrieval. In *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 279–280, Berkeley, USA.

Stolcke, A. (2002). Srilm-an extensible language modeling toolkit. In *The International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 257–286, Denver, USA. ISCA.

Suendermann, D., Evanini, K., Liscombe, J., Hunter, P., Dayanidhi, K., and Pieraccini, R. (2009a). From rule-based to statistical grammars : Continuous improvement of large-scale spoken dialog systems. In *The International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4713–4716, Taipei, Taiwan. IEEE.

Suendermann, D., Liscombe, J., Dayanidhi, K., and Pieraccini, R. (2009b). Localization of speech recognition in spoken dialog systems : How machine translation can make our lives easier. In *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 1475–1478, Brighton, UK. ISCA.

Sutton, S., Novick, D., Cole, R., Vermeulen, P., de Villiers, J., Schalkwyk, J., and Fanty, M. (1996). Building 10,000 spoken dialogue systems. In *The International Conference on Spoken Language Processing (ICSLP)*, volume 2, pages 709–712, Philadelphia, USA. ISCA.

Thomson, B. and Young, S. (2010). Bayesian update of dialogue state : A pomdp framework for spoken dialogue systems. *Compututional Speech Language*, 24(4) :562–588.

Tillmann, C. and Ney, H. (2003). Word reordering and a dynamic programming beam search algorithm for statistical machine translation. *Computational Linguistics*, 29(1) :97–133.

Tur, G., Hakkani-Tur, D., and Schapire, R. (2005). Combining active and semi-supervised learning for spoken language understanding. *Speech Communication*, 45(2) :171–186.

Tur, G., Rahim, M., and Hakkani-Tur, D. (2003). Active labeling for spoken language understanding. In *The European Conference on Speech Communication and Technology (EUROSPEECH)*, Geneva, Switzerland. ISCA.

Tür, G., Wright, J. H., Gorin, A. L., Riccardi, G., and Hakkani-Tür, D. Z. (2002). Improving spoken language understanding using word confusion networks. In *The Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 1137–1140, Denver, USA. ISCA.

Turian, J. P., Wellington, B., and Melamed, I. D. (2006). Scalable discriminative learning for natural language parsing and translation. In Schölkopf, B., Platt, J. C., and Hoffman, T., editors, *Annual Conference on Neural Information Processing Systems Neural Information Processing Systems (NIPS)*, pages 1409–1416, Vancouver, Canada. MIT Press.

Vapnik, N. V. (1982). *Estimation of the Dependences based on empirical data*. Springer.

Vauquois, B. (1968). A survey of formal grammars and algorithms for recognition and transformation in machine translation. In *The International Federation for Information Processing Conference*, pages 254–260, Edinburgh.

Villaneau, J., Antoine, J.-Y., and Ridoux, O. (2004). Logical approach to natural language understanding in a spoken dialogue system. In Sojka, P., Kopecek, I., and Pala, K., editors, *The International Conference on Text, Speech and Dialogue (TSD)*, volume 3206 of *Lecture Notes in Computer Science*, pages 637–644. Springer.

Vogel, S., Ney, H., and Tillmann, C. (1996). HMM-based word alignment in statistical translation. In *The International Conference on Computational Linguistics (COLING)*, pages 836–841, Copenhagen, Denmark.

Walker, M. A., Stent, A., Mairesse, F., and Prasad, R. (2007). Individual and domain adaptation in sentence planning for dialogue. *Journal of Artificial Intelligence Research (JAIR)*, 30 :413–456.

Wang, Y. and Acero, A. (2006). Discriminative models for spoken language understanding. In *The International Conference on Spoken Language Processing (ICSLP)*, pages 61–64, Pittsburgh, USA. ISCA.

Williams, J. D. (2008). Integrating expert knowledge into pomdp optimization for spoken dialog systems. Technical report, The Workshop on Advancements in POMDP Solvers, Chicago, USA.

Woodl, P. C., Hain, T., Johnson, S. E., Niesler, T. R., Tuerk, A., Whittaker, E. W. D., and Young, S. J. (1998). The 1997 HTK broadcast news transcription system. In *DARPA Broadcast News Transcription and Understanding Workshop*, virginia, USA.

Woods, W. A. (1970). Transition network grammars for natural language analysis. *Communications of the ACM*, 13(10) :591–606.

Yvon, F., Zweig, G., and Saon, G. (2004). Arc minimization in finite-state decoding graphs with cross-word acoustic context. *Computer Speech and Language*, 18(4) :397–415.

Zaslavskiy, M., Dymetman, M., and Cancedda, N. (2009). Phrase-based statistical machine translation as a traveling salesman problem. In *The Annual Meeting of the Association for Computational Linguistic (ACL)*, pages 333–341, Suntec, Singapore. ACL.

Zens, R., Och, F. J., and Ney, H. (2002). Phrase-based statistical machine translation. In *Annual German Conference on Advances in Artificial Intelligence*, KI '02, pages 18–32, London, UK. Springer.

Zhou, L., Lin, C.-Y., and Hovy, E. (2006). Re-evaluating machine translation results with paraphrase support. In *The Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 77–84, Sydney, Australia. ACL.

Zue, V., Glass, J., Goddeau, D., Goodine, D., Hirschman, L., Phillips, M., Polifroni, J., and Seneff, S. (1992). The mit atis system : February 1992 progress report. In *The Workshop on Speech and Natural Language*, HLT '91, pages 84–88, Harriman, New York. ACL.

Bibliographie personelle

B. Jabaian, L. Besacier, F. Lefèvre, Investigating multiple approaches for SLU portability to a new language, Dans les actes de *INTERSPEECH (2010)*.

B. Jabaian, L. Besacier, F. Lefèvre, Combination of stochastic understanding and machine translation system for language portability of dialogue systems, Dans les actes de *ICASSP (2011)*,

B. Jabaian, L. Besacier, F. Lefèvre, Comparaison et combinaison d'approches pour la portabilité vers une nouvelle langue dun système de compréhension de loral, Dans les actes de *TALN (2011)* (Papier Long).

F. Lefèvre, D. Mostefa, L. Besacier, Y. Esteve, M. Quignard, N. Camelin, B. Favre, B. Jabaian, L. Rojas-Barahona, Robustness and adaptation of spoken language understanding systems among languages and domains : the PORTMEDIA project, Dans les actes de *LREC (2012)*.

F. Lefèvre, D. Mostefa, L. Besacier, Y. Esteve, M. Quignard, N. Camelin, B. Favre, B. Jabaian, L. Rojas-Barahona, Robustesse et portabilités multilingue et multi-domaines des systèmes de compréhension de la parole : le projet PortMedia, Dans les actes des *JEP (2012)*.

B. Jabaian, F. Lefèvre, L. Besacier, Portability of semantic annotations for fast development of dialogue corpora, Dans les actes des *INTERSPEECH (2012)*.

B. Jabaian, L. Besacier, F. Lefèvre, Comparaison and combination of several robust approches for language portability of a spoken understanding system, En révision pour *IEEE transaction on audio speech and language processing* (Accepté avec revision mineure).

Annexe1 : Expérience comparable de portabilité vers un nouveau domaine

Dans une expérience parallèle, nous avons proposé de porter un corpus existant afin d'obtenir rapidement un nouveau corpus dans la même langue mais pour un domaine different. Pour cela nous proposons d'utiliser le corpus MEDIA pour effectuer une pré-annotation automatique d'un nouveau corpus français. Le nouveau corpus (nommé PM-DOM) concerne le domaine de réservation de billets de théâtre pour le festival d'Avignon. Ce nouveau corpus partage certains concepts avec le corpus MEDIA tel que les dates et les lieux, mais aussi il a ses propres concepts représentant la spécificité du domaine (acteur, pièce, ...).

Le modèle utilisé pour la pré-annotation a été appris directement sur les données MEDIA (de la même langue) et un étiqueteur d'entités nommées a été ajouté pour prendre en compte les nouveautés du domaine. Cet étiqueteur est basé sur une liste d'entités nommées du nouveau domaine. Cette liste est composée de noms d'auteurs, de noms de pièces et de lieux extraits du programme du festival d'Avignon. La pré-annotation est une combinaison entre l'étiqueteur sémantique apprit sur MEDIA et la sortie de l'étiqueteur d'entités nommées.

Nous avons aussi appris un modèle sur les données PM-DOM afin d'évaluer ce corpus. En parallèle, nous avons évalué le corpus de référence utilisé pour la pré-annotation (domaine touristique) sur l'ensemble de test PM-DOM (réservation de places). Les résultats sont présentés dans le tableau 8.1.

Le modèle PM-DOM donne un CER de 19,1%, ce qui est assez bon relativement à la taille de données d'apprentissage. Le modèle MEDIA utilisé pour la pré-annotation (appris sur le corpus de base) donne une CER de 41,2% sur le corpus PM-DOM. Ce faible score est dû au fait qu'il y a une différence considérable dans les concepts entre le corpus MEDIA et le corpus PM-DOM.

Cette différence de concepts entre les deux corpus a une influence significative sur la combinaison directe des deux corpus. Contrairement à ce qu'on a montré pour la nouvelle langue, une concaténation des corpus MEDIA et PM-DOM n'est pas évidente. Comme le montre le tableau 8.1, le modèle combiné donne un CER de 37,5%, ce qui est beaucoup moins bon que le modèle appris sur PM-DOM seul. Cette dégradation provient du fait que la combinaison conduit à un nombre important d'insertions dans

Modèle	Test	Sub	Del	Ins	CER
PM-DOM		3,2	13,0	2,9	19,1
MEDIA		11,2	25,6	4,4	41,2
PM-DOM + MEDIA	PM-DOM	5,8	2,8	28,9	37,5
PM-DOM + F-MEDIA		3,0	13,2	3,3	19,5
PM-DOM + F'-MEDIA		3,4	11,8	3,9	18,9

TABLE *8.1* – *Evaluation (CER %) des modèles français sur le test* PM-DOM.

les hypothèses de sortie. Ces insertions sont des concepts du corpus MEDIA n'existant pas dans le corpus PM-DOM. Pour obtenir une combinaison plus efficace, nous proposons de filtrer le corpus MEDIA, en éliminant les concepts du corpus MEDIA qui n'existent pas dans le corpus PM-DOM.

Nous considérons deux options de filtrage. La première consiste à éliminer tous les segments annotés avec des concepts indésirables (pas utilisés dans le corpus PM-DOM) et nous obtenons un corpus MEDIA filtré dénommé F-MEDIA (10K phrases vs 13K de MEDIA). L'autre option de filtrage consiste à éliminer entièrement les phrases qui contiennent des concepts indésirables (corpus F'-MEDIA, 7K phrases). Chacun de ces corpus filtrés est ensuite concaténé avec le corpus PM-DOM et un modèle est appris sur le corpus combiné.

Il est important de mentionner que le corpus F-MEDIA contenait plus de phrases que F'-MEDIA, mais certaines de ces phrases sont bruitées. Ce bruit provient de la suppression des segments indésirables apparaissant à l'intérieur de la phrase. Il y a une hypothèse forte derrière cette stratégie qui est que la suppression d'un segment dans une phrase, mènera généralement à un autre énoncé acceptable. Mais, évidemment, la nature précise des concepts enlevés a un grand impact sur le résultat réel et de nombreuses phrases peuvent être finalement incorrectes.

Les résultats présentés dans le tableau 8.1 montrent que le modèle appris sur la combinaison de PM-DOM avec F'-MEDIA (18,9%) est plus performant que celui appris sur PM-DOM avec F-MEDIA (19,5%). La combinaison est meilleure avec moins de données qu'avec plus de données bruitées. Cette combinaison augmente très légèrement les performances de notre modèle par rapport au modèle PM-DOM (19,1% vs 18,9%).